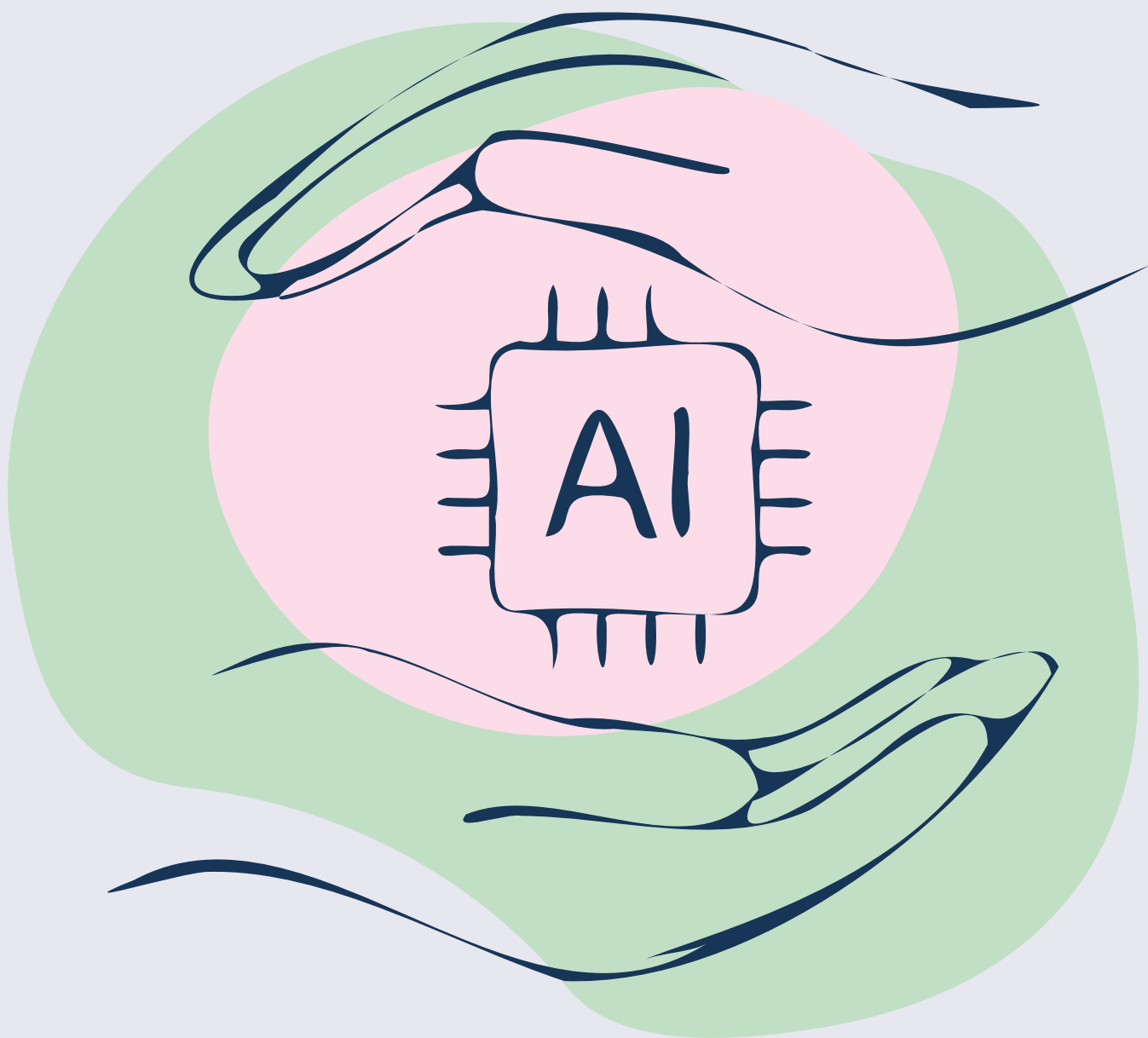


Tekoäly ja sukupuolten tasa-arvo



Tane

Tasa-arvoasiain
neuvottelukunta

Tasa-arvoasiain neuvottelukunta (Tane)

Tekoäly ja sukupuolten tasa-arvo

Pekka Tuominen ja Karoliina Snell

tane.fi

Julkaisusarja: TANE-julkaisuja 2025:24

Julkaisija: Tasa-arvoasiain neuvottelukunta

Tekijät: Pekka Tuominen ja Karoliina Snell

Kansikuva: Duran Creative Oy

Taitto: Duran Creative Oy

Asiasanat: tekoäly; tasa-arvo; työelämä; lainsäädäntö

ISBN 978-952-00-4408-4

ISSN 1798-1107



Tekoäly ja sukupuolten tasa-arvo

Selvitys tarjoaa ajankohtaisen katsauksen tekoälyn kehityksen ja sukupuolten tasa-arvon yhtymäkohtiin. Se tuo esiin, millä tavalla yhteiskunnan eriarvoisuudet peilautuvat tekoälyyn: jos yhteiskunnassa on puutteita tasa-arvon toteutumisessa, näkyy tämä todennäköisesti myös tekoälysovelluksissa. Selvitys tarkastelee kriittisesti tekoälyn vaikutuksia sukupuolten tasa-arvoon kotimaisten ja kansainvälisten esimerkkien kautta. Lisäksi se avaa perus- ja ihmisoikeuksiin, tasa-arvoon ja tekoälyyn liittyvää sääntelykehikkoa sekä tuo esiin tekoälyn kehitykseen ja työelämään kytkeytyviä sukupuolittuneita kysymyksiä.

Selvityksessä kuvataan, kuinka tekoälyjärjestelmien suunnittelun ja järjestelmien tekemien päätelmien pohjana olevan datan taustalla on monenlaista ihmisten tekemää luokittelua ja datatyötä. Datan vinoumat ja huonosti kehitetyt ja hallitut tekoälyjärjestelmät voivat ylläpitää sukupuolistereotyyppioita, aiheuttaa syrjiviä käytäntöjä tai loukata ihmisten perusoikeuksia.

Tekoälyn vaikutuksia sukupuolten tasa-arvoon ja syrjintään on vaikea täysin ennakoida. Se, missä ja miten tekoälyä käytetään, on kuitenkin ihmisten päätösten seurausta. Vaikka tekoäly voi ylläpitää ja jopa vahvistaa sukupuoleen perustuvaa eriarvoisuutta, haitallisia vaikutuksia voidaan ehkäistä huolellisella vaikutusten arvioinnilla sekä perehtymällä käyttökontekstiin ja tekoälyn toimintalogiikkaan.

Asiasanat: tekoäly; tasa-arvo; työelämä; lainsäädäntö

AI och jämställdhet

Rapporten ger en aktuell översikt över beröringspunkter mellan utvecklingen av artificiell intelligens (AI) och jämställdhet. Den belyser hur ojämlikheter i samhället återspeglas i AI: om jämlikheten i samhället är bristfällig kommer detta sannolikt att återspeglas i AI-applikationer. Rapporten granskar kritiskt hur AI påverkar jämställdheten mellan könen genom nationella och internationella exempel. Den öppnar också upp regelverket för de grundläggande och mänskliga rättigheterna, jämställdhet och AI samt belyser genusfrågor i samband med AI-utveckling och arbetslivet.

I rapporten beskrivs hur de data som ligger till grund för utformningen av AI-system och de slutsatser de drar bygger på klassificering av olika slag och dataarbete som utförs av människor. Snedvridning av data och dåligt utvecklade och hanterade AI-system kan upprätthålla könsstereotyper, leda till diskriminerande praxis eller kränka människors grundläggande rättigheter.

AI:s inverkan på jämställdhet och diskriminering är svår att förutse helt. Var och hur AI används är dock ett resultat av mänskliga beslut. Även om AI kan upprätthålla och till och med förstärka ojämlikhet, kan negativa effekter förebyggas genom noggranna konsekvensbedömningar och genom att förstå i vilka sammanhang AI används och enligt vilken logik den fungerar.

Nyckelord: artificiell intelligens; jämställdhet; arbetsliv; lagstiftning

AI and gender equality

The report provides an up-to-date overview of the interfaces between gender equality and the development of artificial intelligence (AI). It highlights how inequalities in society are reflected in AI: if there are shortcomings in the realisation of gender equality in society, this is also likely to be reflected in AI applications. The report takes a critical look at the impacts of AI on gender equality through national and international examples. It also discusses the regulatory framework related to fundamental and human rights, gender equality and AI and highlights gendered issues associated with AI development and working life.

The report describes how the data underpinning both the planning of AI systems and the conclusions drawn by these systems is based on various classifications and data work made by people. Biases in data and poorly developed and managed AI systems can maintain gender stereotypes, lead to discriminatory practices or violate people's fundamental rights.

It is difficult to fully predict the impacts of AI on gender equality and discrimination. However, it is people who decide where and how AI is used. Although AI can maintain and even reinforce gender inequalities, its adverse effects can be prevented by assessing its impacts carefully and by examining the context of use and the operating logic of AI.

Keywords: artificial intelligence; gender equality; working life; legislation

Sisällysluettelo

1. Johdanto	7
1.1 Selvityksen tarkoitus ja tausta	7
1.2 Johdatus tekoälyyn ja tekoälyn kehityksen ihmisvaikutuksiin	9
1.3 Tekoälyn keskeiset riskit	12
1.4 Selvityksessä käytettävät käsitteet	13
2. Tekoälyn toisintamat sukupuolistereotyytiat ja syrjintä	16
2.1 Datan vinoumat, puutteet ja luokittelu	16
2.2 Miksi data ei ole koskaan neutraalia?	19
2.3 Algoritminen syrjintä ja stereotyypioiden ylläpito	20
2.4 Generatiivinen tekoäly ja syrjintä	23
2.5 Ennustavan tekoälyn haasteet	25
2.6 Tekoäly ylläpitää, uusintaa ja vahvistaa yhteiskunnan eriarvoisuuksia	26
3. Tekoälyn kehityksen vaikutus työelämän muutokseen	29
3.1 Tekoälyteollisuuden sukupuolijakauma	29
3.2 Vinoumat rekrytoinnissa	30
3.3 Tekoäly muuttaa työelämää	30
4. Kansainvälinen ja kansallinen tekoälyyn liittyvä sääntely	33
4.1 Sukupuolten tasa-arvoon liittyvä sääntely tekoälysääntelyn taustalla	33
4.2 Tekoälysäädös ja muu kansainvälinen sääntely	34
4.3 Sääntelyn haasteet ja mahdollisuudet	37
5. Johtopäätökset	41
Lähteet	43

1. Johdanto

1.1. Selvityksen tarkoitus ja tausta

Tämä selvitys tarjoaa katsauksen tekoälyn kehityksen ja sukupuolten tasa-arvon yhtymäkohtiin. Sen tavoitteena on lisätä tietoa siitä, miten sukupuoli ja tasa-arvo liittyvät tekoälyn kehittämiseen ja sovelluksiin. Lisäksi selvitys lisää tietoa sukupuolten tasa-arvoon ja tekoälyyn liittyvistä mahdollisuuksista ja uhista. Selvityksessä keskitytään tekoälyn käytön tuottamiin ja ylläpitämiin sukupuolisen syrjinnän muotoihin, niitä tukeviin sukupuolistereotypioihin sekä mahdollisuuksiin edistää sukupuolten tasa-arvoa sääntelyn keinoin.

Selvityksen ovat kirjoittaneet tasa-arvoasiain neuvottelukunnan toimeksiannosta yliopistotutkija VTT Pekka Tuominen ja ohjelmajohtaja VTT Karoliina Snell Helsingin yliopistosta yhteistyössä SHIELD-ohjelman kanssa. SHIELD eli ”Turvallisuus ja luottamus algoritmien aikakaudella” on laaja monitieteinen tutkimusohjelma, jota rahoittaa Suomen Akatemian alainen Strategisen tutkimuksen neuvosto vuosina 2022–2028.

Selvitys pohjaa ajankohtaiseen tutkimuskirjallisuuteen sekä kotimaiseen ja kansainväliseen keskusteluun. Selvityksessä on käytetty myös media-aineistoja ja taustoittavia keskusteluja eri alojen tutkijoiden kanssa. Selvityksen tueksi järjestettiin kaksi pyöreän pöydän keskustelua, joissa yhteensä 17 tekoälyn ja sukupuolten tasa-arvon asiantuntijaa eri aloilta kokoontuivat keskustelemaan selvityksen teemoista oman tutkimuksensa ja ammattiosaamisensa valossa.

Tekoälyä käytetään monilla eri aloilla, ja uusia ratkaisuja kehitetään nopealla tahdilla. Nopeasti muuttuvan luonteensa vuoksi yhdessä selvityksessä ei ole mahdollista eikä edes tarkoituksenmukaista pyrkiä tarkastelemaan kaikkia mahdollisia tekoälyn soveltamiseen liittyviä vaikutuksia sukupuolten tasa-arvon näkökulmasta. Tässä selvityksessä kuvailtujen esimerkkien avulla on kuitenkin mahdollista ymmärtää, miten ja miksi tekoälyjärjestelmät voivat aiheuttaa vinoumia ja syrjintää, ja miksi aiheeseen on tärkeää kiinnittää aktiivisesti huomiota.

Tekoälyn kehittämistä ja käyttöä ohjaavat ihmiset sekä yhteiskunnassa vallalla olevat arvot ja toimintatavat, mikä käy ilmi aiemmista tapaustutkimuksista ja tässä selvityksessä esiin nostettavista esimerkeistä. Yhteiskunnan eriarvoisuudet peilautuvat tekoälyyn: jos yhteiskunnassa on puutteita tasa-arvon toteutumisessa, näkyy tämä todennäköisesti myös tekoälysovelluksissa. Sukupuoleen liittyviä stereotypioita tai tekoälyn opettamiseen käytettyä vinoutunutta dataa on vaikea korjata vain itse tekoälysovelluksissa, jos sovellusten kehittäjien ajattelu, opetusdata tai sovelluksen käyttötarkoitus eivät huomioi

sukupuolittuneita eroja ja valtasuhteita. Virheet voivat johtua väärän teknologian valinnasta, sen väärästä käyttötavasta, huolimattomasta testauksesta, ilmiön huonosta tuntemisesta tai siitä, että tekoälyn käytön kaikkia vaikutuksia ei ole pohdittu tarpeeksi. Saattaa myös olla, että tekoälyä käytetään kokonaan väärin tehtäviin¹.

Tekoälyteknologian kehittäminen sukupuolten tasa-arvon näkökulmasta oikeudenmukaisempaan suuntaan on siis yhteiskunnallinen kysymys, jossa teknologialla, politiikalla, lainsäädännöllä sekä sovellusten kehittäjillä ja käyttäjillä on kaikilla tärkeä tehtävä tasa-arvon edistäjinä. Selvitys sisältää tärkeitä huomioita pohdittavaksi sekä julkisella että yksityisellä sektorilla, joissa molemmissa parhaillaan keskustellaan siitä, miten tekoälyjärjestelmiä voidaan kehittää ja ottaa käyttöön vastuullisesti. Tekoälyn käyttö syrjimättömällä tavalla perustuu tietoiisiin sosiaalisiin ja poliittisiin valintoihin, jotka heijastavat laajempia yhteiskunnallisia arvoja². Sääntelyn tai tekoälyä koskevan politiikan kohteena ei voi olla tekoäly irrotettuna laajemmista yhteiskunnallisista ilmiöistä, vaan siinä on tärkeää huomioida vuorovaikutus teknologian ja yhteiskunnan välillä³. Tekoäly voi tuoda mukanaan tuottavuuden kasvua, mutta ainoastaan, jos samalla kehitetään kaikkien kansalaisten kyvykkyyksiä käyttää sitä. Jos teknologiat syrjivät sukupuolen perusteella ja käyttävät sukupuolen mukaan vinoutuneita aineistoja, nopeasti leviävä tekoälyn käyttö yhä uusilla alueilla voi heikentää sukupuolten tasa-arvoa. Sama koskee tietysti myös muita syrjintäperusteita, kuten ikää, etnisyyttä tai esimerkiksi vammaisuutta. Onkin tärkeää, että tekoälyn ongelmiin tasa-arvon ja yhdenvertaisuuden osalta tartutaan aktiivisesti kaikkien osapuolten toimesta: tekoälyn kehittäjien, käyttäjien ja ennen kaikkea asiasta päättävien taholta.⁴

Selvityksen rakenne

Selvitys alkaa katsauksella erilaisiin tilanteisiin, joissa tekoäly ja sukupuolten tasa-arvo ovat kytkeytyneet toisiinsa. Selvityksen ensimmäisessä luvussa kerrotaan datafikaatiosta – digitaalisessa muodossa olevan informaation kiihtyvistä keräämisestä – ja tekoälyn leviämisestä yhä useammalle elämän alueelle. Tämän lisäksi katsauksessa esitellään lyhyesti tekoälyn käytön riskit perusoikeuksille ja syrjimättömyydelle sekä selvityksessä käytettävät keskeiset termit.

Selvityksen toinen luku keskittyy tekoälyjärjestelmissä esiintyviin vinoumiin ja vääristymiin sekä niiden toisintamiin sukupuolistereotyyppioihin ja syrjiviin mekanismeihin. Syrjintää käsitellään ja kuvataan useiden esimerkkien avulla. Luvun lopussa esitetään, kuinka sekä vääristymiä korjaamalla että syrjiviin rakenteisiin puuttumalla voidaan edistää tasa-arvoa tekoälyn käytössä.

Kolmannessa luvussa tarkastellaan tekoälyn ja sukupuoleen perustuvan syrjinnän suhdetta työelämässä. Luvussa tarkastellaan naisten vähäistä määrää teknologia-alalla, tekoälyn käyttöä rekrytoinnissa sekä sitä, kuinka tekoäly muuttaa työelämää.

¹ Toivonen, 2023, s. 45

² Bartoletti & Xenides, 2023, s. 72

³ Bartoletti & Xenides, 2023, s. 14

⁴ UNESCO, OECD, IDB, 2022

Neljäs luku keskittyy tekoälyn sääntelyyn. Luvussa käsitellään sääntelyn periaatteita, niiden suhdetta perusoikeuksiin ja sukupuolten tasa-arvoon sekä keinoja ulottaa sääntely eri tavoin rajattuihin ympäristöihin. Luvun loppupuolella kuvataan, kuinka sääntelyn keskeiset periaatteet nousevat usein data-aktivistien keskusteluista ennen kuin niitä viedään kansalliseen ja kansainväliseen päätöksentekoon. Lopuksi esitellään selvityksen keskeiset tulokset ja johtopäätökset.

1.2. Johdatus tekoälyyn ja tekoälyn kehityksen ihmisvaikutuksiin

Tekoälysovellusten nopea kehittyminen on herättänyt keskustelua niin politiikassa, mediassa kuin tutkimuksenkin parissa. Julkisuudessa ja aiemmassa kirjallisuudessa on käsitelty tapauksia, joissa tekoälysovellukset ovat tuottaneet epätasa-arvoisia tuloksia. Media on esimerkiksi nostanut esiin, kuinka koneoppimisella automatisoitu rekrytointi tarjosi naishakijoille pienempiä palkkoja kuin miehille ja törmäystestisimulaatiot liikenteessä mallinnettiin tekoälyllä miesten fysiologian pohjalta ja niiden mukaan rakennetut autot olivat naisille vaarallisempia.

Edellä mainitun kaltaiset uutiset ja tapaukset syrjivistä käytännöistä paljastavat tekoälyn käyttöön liittyviä haasteita, mutta antavat harvemmin neuvoja, kuinka niihin voi puuttua. Ratkaisut tekoälyn tuottamiin syrjiviin käytäntöihin eivät mahdu yhteen muottiin – ongelmat eivät ratkea vain algoritmeja säätämällä, vaan ne vaativat monimutkaisten yhteiskunnallisten rakenteiden ymmärtämistä, muovaamista ja demokraattista osallistumista. Sukupuolten tasa-arvo ei myöskään pelkisty yksinkertaisiin jakoihin, vaan tasa-arvoa tulee lähestyä useiden risteävien tekijöiden summana, identiteetin osina, jotka ovat sidoksissa aikaansa ja muuttuvat sen mukana.

Varsinkin tahattoman syrjinnän ja epätasa-arvoisten käytäntöjen huomaaminen ja korjaaminen tekoälyjärjestelmissä on vaikeaa. Syrjiviä käytäntöjä syntyy usein koneoppimisen sivutuotteena, vaikka niitä yritettäisiin tietoisesti välttää. Eräs kirjallisuudessa usein kuvattu esimerkki käsittelee tietokoneohjelman tekemää arviota henkilön luottoriskistä asuinpaikan perusteella. Jos asukkaat jollakin tietyllä asuinalueella hoitavat tilastollisesti lainansa huonosti ja tätä tietoa sovelletaan yksittäisen asiakkaan lainahakemuksen arviointiin, kyseessä on ennakkoluuloon perustuva syrjintä, ei todellisen lainanottokyvyn arviointi⁵.

Toisessa tunnetussa esimerkissä internet-kuvahaku sukupuolittaa tuloksia tavalla, joka vahvistaa olemassa olevia sukupuolistereotyyppioita: haut lääkäreistä, johtajista ja muista arvostetuista ammattiteistä tuottavat useimmin kuvia valkoisista miehistä, kun taas naisten kuvia alkaa ilmaantua matalammin palkattujen ammattinimikkeiden takaa. Jotkut teknologiayritykset ovat pyrkineet korjaamaan ongelmaa lisäämällä naisten ja rodullistettujen ihmisryhmien osuutta tuloksissa korostamalla näitä tekoälysovelluksissaan. Sovellusten toimintaperiaatteet on kuitenkin jätetty ennalleen ja ne voivat uusintaa syrjiviä rakenteita muissa hauissa. Reagoinnissa syrjivään tekoälyyn ei siis riitä, että ohjelmien ja sovellusten virheitä ja vääristymiä korjataan yksi hakukriteeri kerrallaan⁶.

⁵ Toivonen, 2023

⁶ Gangadharan & Niklas, 2019; Dencik, 2020

Tekoälyn käyttö ja soveltaminen yhteiskunnassa nojaa ihmisten tekemiin päätöksiin, ei asioiden vääjäämättömään etenemiseen. Tekoälyn kietoutumisessa arkeemme ei myöskään ole kyse jostakin aivan uudesta tilanteesta, joka noudattaa omia, meille vielä tuntemattomia lakejaan. Enemminkin kyse on yhteiskunnallisten ja teknologisten muutosten jatkumosta – yhteiskunnan digitalisoitumisesta, internetin ja sosiaalisen median yhä laajemmasta leviämisestä osaksi arkea sekä suuria datamassoja (big data) hyödyntävistä sovelluksista, joita käytetään osana päätöksentekoa. Tekoälyssä ei ole kysymys vain teknologisista ratkaisuista, joiden tarkoituksena on parantaa tai tehostaa ihmisten elämänlaatua ja kommunikaatiota tai yritysten tuottavuutta ja kehittymistä, vaan kokonaisvaltaisista yhteiskunnallisista kehityskuluista, joita edistetään ja vastustetaan monilla eri tavoilla.

Huomattava osa tekoälyä koskevasta keskustelusta rajautuu joko teknologisiin ratkaisuihin tai niistä erotettuihin yhteiskunnallisten ilmiöiden analyysiin. Näitä kahta osa-aluetta ei kuitenkaan tulisi nähdä toisistaan erillisinä. Tekoälyn yhteydessä käytetään usein termiä datafikaatio kuvaamaan kehitystä, jossa elämän kaikista mahdollisista puolista kerätään yhä kiihtyvään tahtiin dataa, digitaalisessa muodossa olevaa informaatiota, jonka perusteella tehdään yhteiskuntaa ja ympäristöä muokkaavia päätöksiä⁷. Ihmisistä on aina kerätty tietoa eri tarkoituksiin, mutta digitaaliset teknologiat ovat antaneet uusia työkaluja tehdä tätä entistä nopeammin ja laajemmin. Data ja tekoäly liittyvät kiinteästi yhteen. Koneoppimisessa tarvitaan opetusdataa, eli esimerkiksi tekstiä, numeroita tai kuvaa, jonka perusteella ohjelma oppii tekemään päätelmiä. Tämän jälkeen päätelmiä voidaan sitten soveltaa toisiin data-aineistoihin.

Yhä useammalle elämänalueelle leviävä datafikaatio vaikuttaa moninaisilla tavoilla arkeen ja tuottaa uusia risteäviä eriarvoisuuden muotoja. Datafikaatioon liittyvät valta-asetat ovat läsnä kaikissa tekoälyn käytön vaiheissa: kehitystyössä käytetyistä aineistoista, niiden perusteella laadituista sovelluksista ja tuotteista sekä ympäristöissä, joissa niitä käytetään⁸. Keskittyminen pelkästään joko teknologiaan tai yhteiskuntaan rajoittaa analyysiä ja ohjaa sen väärille raiteille.

Miksi tekoälyä ja sukupuolten tasa-arvoa on tärkeää tarkastella yhdessä?

Tekoälyä ja sukupuolten tasa-arvoa tulee tarkastella yhdessä, koska niiden yhteisvaikutukset ulottuvat yhä kasvavalle osalle elämänalueita. Tekoälyä käytetään sekä yksityisellä että julkisella sektorilla ja monilla soveltamisaloilla, aina työnhausta sosiaaliseen mediaan ja sosiaali- ja terveystalviteista rikosten torjuntaan.

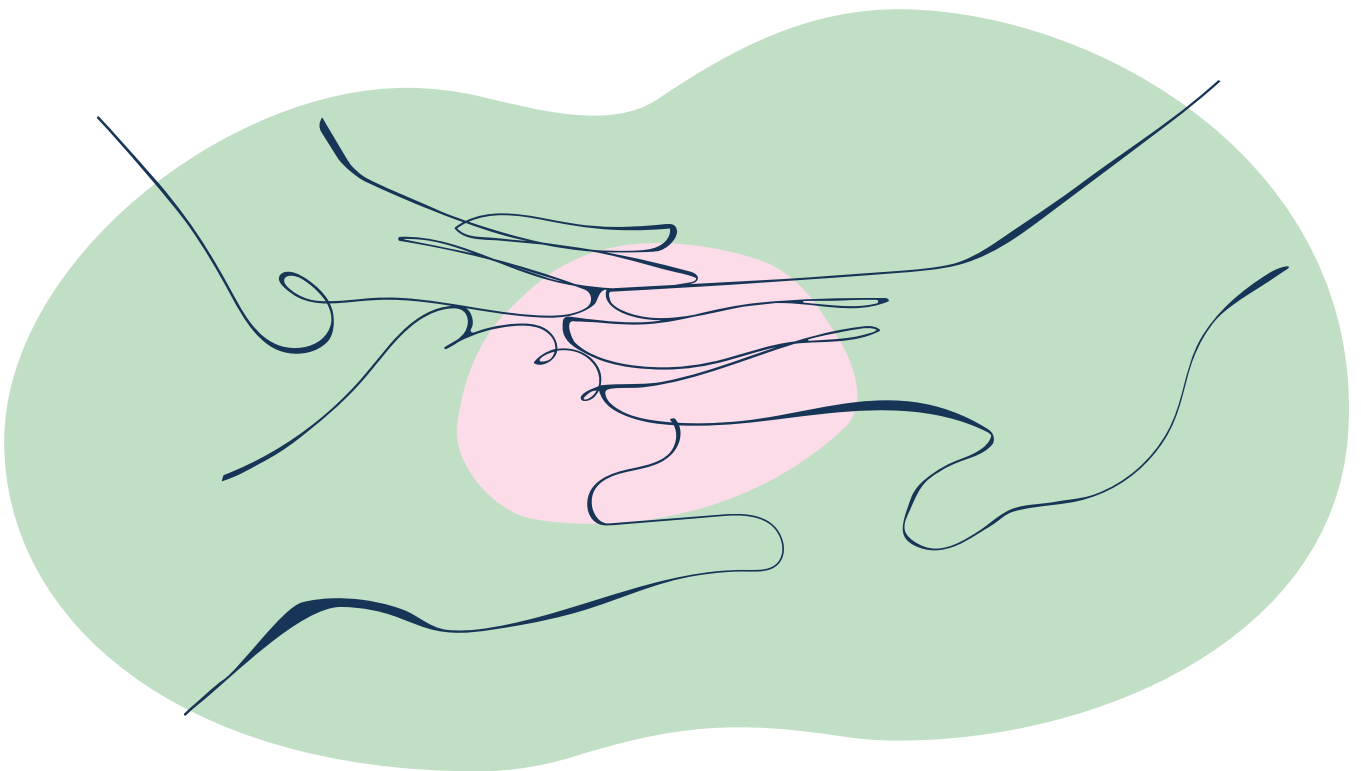
Tekoälyjärjestelmien kehittämiseen ja käyttöön liittyy monenlaisia perusoikeuskysymyksiä riippumatta siitä, käytetäänkö tekoälyä yksityisellä vai julkisella sektorilla. Tekoälyjärjestelmien keskeinen tarkoitus on usein luokitella ja tehdä eroja erilaisten ilmiöiden, muuttujien tai ihmisten välillä. Tällöin voidaan määritellä, ketkä ovat esimerkiksi jonkun palvelun tai hoidon tarpeessa tai minkälaiset ihmiset ovat kiinnostuneita jostain tuotteesta. Luokittelun kautta tekoälyjärjestelmät voivat kuitenkin myös juurruttaa, ylläpitää ja

⁷ Hoeyer, 2023

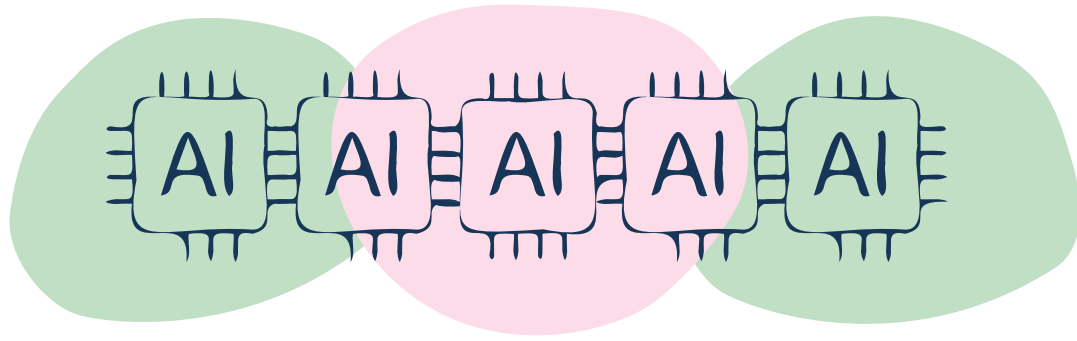
⁸ D'Ignazio & Klein, 2023, s. 201

vahvistaa eriarvoisuutta. Huonosti kehitetyt ja hallitut tekoälyjärjestelmät voivat loukata ihmisten perusoikeuksia monin tavoin. Sukupuolten tasa-arvon kohdalla keskeisiä esimerkkejä ovat tekoälysovellusten tuottamat sukupuoliin liittyvät stereotypit, kuten vain esimerkiksi miehille toimivat algoritmit ja vinoutunut data. Tekoäly mahdollistaa myös uusin tavoin seksuaalista ja sukupuoleen kohdistuvaa häirintää. Tekoälyllä voidaan esimerkiksi automatisoida loukkaavien viestien lähettäminen, laajentaa häirinnän vaikutuspiiriä ja tehdä siitä anonyymimpää. Syrjivien mekanismien toimintaa avataan selvityksessä myöhemmin esimerkkien avulla.

Syrjivän tekoälyn vaikutukset yhteiskuntaan ovat haitallisia yksilöiden, demokratian ja talouden näkökulmasta. Esimerkiksi verkkoväkivalta ja vainoaminen digitaalisilla alustoilla ovat kasvavan julkisen huomion aiheita. On tärkeää pohtia, mistä tekoälyyn liittyvät tasa-arvo-ongelmat kumpuavat: itse tietokoneohjelmilla ei tietenkään ole sukupuolta tai muita sen kaltaisia ihmisten ominaisuuksia. Tästä huolimatta tekoälyn soveltaminen tuotteissa, palveluissa ja päätöksenteossa heijastelee sen toteuttajien asenteita ja maailmankuvaa. Tällä hetkellä huomattava osa suurten teknologiayritysten johdosta on valkoisia miehiä globaalista pohjoisesta. Naisten ja rodullistettujen ihmisten puuttuminen johtopaikoilta vaikuttaa väistämättä tekoälyratkaisuihin. Tekoälyyn liittyvä syrjintä aiheuttaa monenlaisia ongelmia myös yrityksille. Vuonna 2022 tehty kansainvälinen kysely osoittaa, että yli kolmannes (36 %) yrityksistä kärsi liiketoiminnassaan tekoälyn aiheuttamista vääristymistä. Ongelma johti menetettyihin tuloihin, työntekijöihin ja asiakkaisiin, oikeudenkäyntikuluihin ja monenlaisiin mainehaittoihin⁹. On tärkeää ymmärtää, että sukupuolten tasa-arvolle sensitiivisempi ja siihen liittyvät ongelmat huomioiva tekoälyn käyttö on sekä kilpailuetu että yhteiskunnallista luottamusta kohentava suunta.



⁹ Bartoletti & Xenides, 2023, s. 10



1.3. Tekoälyn keskeiset riskit

Tekoälyn käyttö on monilla aloilla vasta lapsenkengissään ja sille asetetaan helposti epärealistisia odotuksia. Hannu Toivonen listaa kirjassaan *Mitä tekoäly on? 100 kysymystä ja vastausta*¹⁰ monta negatiivista esimerkkiä, joissa tekoälyjärjestelmät ovat tuottaneet epäoikeudenmukaisia tuloksia ja virheitä, tai ne eivät ole toimineet niin kuin on oletettu:

”Ison-Britannian postilaitos syyllistyi maan historian suurimpaan oikeusmurhaan, kun sen työntekijöitä tuomittiin kavalluksesta Fujitsun tekemän ohjelmointivirheen vuoksi. Vuosien taistelun jälkeen 39 oikeuden tuomiota peruttiin vuonna 2021; erilaisia hyvitysvaatimuksia on tuhansia. IBM:n kehittämä Watson-järjestelmä tuotti syöpäpotilaiden lääkäreille hengenvaarallisia lääkesuosituksia. Microsoftin julkaisema Tay-niminen chatbot kääntyi vuorokaudessa kehumaan Hitleriä, vaikka sen piti nostaa ihmisen ja koneen välinen keskustelu uudelle tasolle. Amazonin kasvontunnistusohjelma ”tunnisti” 28 Yhdysvaltain kongressin edustajaa rikollisiksi. Googlen kuvantunnistusohjelma merkitsi valokuvien mustia ihmisiä gorilloiksi. Applen iPhone:n kasvontunnistusta pystyi huijaamaan naamarilla. Ja niin edelleen.”¹¹

Syitä tekoälyjärjestelmien tuottamiin epäoikeudenmukaisiin tuloksiin ja virheisiin on useita:

- Tekoälyjärjestelmille syötetty opetusdata tai data, jonka pohjalta päätelmät tehdään, on vinoutunutta.
- Tekoälyjärjestelmiä kehitetään kaukana niiden sovelluskohteista. Ihmiset, jotka kehittävät järjestelmiä ja algoritmeja eivät välttämättä tunnista aiheeseen liittyviä vinoumia, stereotyyppioita tai syrjiviä käytänteitä.
- Tekoälyjärjestelmät keskittyvät sellaisten ilmiöiden ja ominaisuuksien analysointiin, joista on saatavilla helposti dataa. Näin ne yksinkertaistavat oletuksia ihmisten toiminnasta, motiiveista ja vuorovaikutuksesta.
- Tekoälyjärjestelmät ehdottavat usein käyttäjille samankaltaisia sisältöjä, jonka takia muunlaiset ajatusmallit ja ratkaisut jäävät sivuun.
- Tekoälyjärjestelmissä käytetään vanhentunutta dataa tulevaisuuden ennustamiseen.
- Tekoälyjärjestelmät saattavat toimia tavoilla, joita niiden suunnittelijat ja käyttäjät eivät pysty ennakoimaan, eikä jälkikäteen ole mahdollista päätellä, miten järjestelmä on päätenyt tulokseen.

Yllä kuvattuja ongelmia sekä tekoälyjärjestelmien luotettavuutta ja syrjimättömyyttä yritetään parantaa ja korjata monella tapaa. Keinot liittyvät niin uusiin teknisiin ratkaisuihin,

¹⁰ Toivonen, 2023

¹¹ Toivonen, 2023, s. 25

lainsäädäntöön ja valvontaan, koulutukseen ja datalukutaidon parantamiseen, kuin myös eettisten periaatteiden noudattamiseen. Tekoölyyn liittyvissä eettisissä keskusteluissa on nostettu esiin periaatteita, joita tekoölyjärjestelmien tulisi noudattaa. Näitä ovat esimerkiksi oikeudenmukaisuus, ihmisen itsemääräämisoikeuden kunnioittaminen ja järjestelmien läpinäkyvyys ja selitettävyys¹². Luvussa 3 käydään tarkemmin läpi keskeisiä tekoölyyn ja sukupuolten tasa-arvoon liittyviä eettisiä ja lainsäädännöllisiä näkökulmia.

Kansalaisten luottamus tekoölyjärjestelmiin on tärkeää, koska niitä käytetään kasvavassa määrin esimerkiksi julkisissa palveluissa. Queenslandin yliopiston ja KPMG:n vuonna 2023 tekemässä kyselyssä tarkasteltiin 17 tekoölyn käytössä johtavan maan kansalaisten asenteita tekoölyä kohtaan. Vaikka suomalaiset suhtautuivat tekoölyyn periaatteessa myönteisesti ja hyväksyivät sen käytön, he luottivat siihen vastaajista vähiten (16 %). 76 % suomalaisista vastaajista oli huolissaan tekoölyyn liittyvistä riskeistä ja 67 % koki sen yhteiskunnalliset vaikutukset arvaamattomina¹³. Arjessa suomalaiset liittävät tekoölyn ja algoritmit useimmiten sosiaaliseen mediaan ja mainostamiseen, eivätkä niinkään omiin talousasioihinsa, ennakoivaan poliisitoimintaan, liikenteeseen tai sosiaalietuuksien jakamiseen. Yhteiskunnat ovat toki myös sisäisesti moninaisia – on helppo kuvitella, että esimerkiksi turvapaikanhakijat Suomessa olisivat korostaneet vastaavassa kyselyssä eri asioita erilaisen elämäntilanteensa vuoksi¹⁴.

Syrjimätön tekoöly on merkittävä kysymys Suomen kaltaisessa avoimessa yhteiskunnassa, jossa yhteiskunnallinen luottamus on suurta – jos se rapautuu suhteessa nopeaan kasvuun tekoölyn käytössä, kehityksellä on varmasti laajoja yhteiskunnallisia vaikutuksia. Tähän haasteeseen täytyy reagoida ennakoivasti, jotta hyvän hallinnon periaate säilyy¹⁵. Vaikka suurinta valtaa tekoölyn kehitystyössä käyttävät globaalit toimijat, tekoölyjärjestelmien vaikutukset ovat erilaisia eri maissa ja ympäristöissä. Sukupuolten tasa-arvoon liittyvät ongelmat, kysymykset ja edistysaskeleet ovat myös hyvin erilaisia eri puolilla maailmaa. Siksi niitä pitää tarkastella systemaattisesti aina kun suunnitellaan tai toteutetaan tekoölyn käyttöön perustuvia järjestelmiä, toimintaperiaatteita tai sääntelymekanismeja¹⁶. Suomen pitää osana Euroopan unionia löytää juuri sille sopiva tie tekoölyn käytössä, aktiivisena ja itsenäisenä toimijana, mutta samalla jatkuvassa globaalissa vuorovaikutuksessa.

1.4. Selvityksessä käytettävät käsitteet

Tekoöly

Tekoölyllä tarkoitetaan kaikkia jossain määrin autonomisesti toimivia järjestelmiä, jotka tuottavat päätelmiä, suosituksia tai sisältöä ohjelmalle asetettujen ohjeiden mukaisesti. Tekoöly käsitteenä kattaa siis hyvin erilaisia menetelmiä ja sovelluksia. Tekoölystä puhutaan kuitenkin usein harhaanjohtavasti, ikään kuin tekoöly olisi yksi suuri ajatteleva kokonaisuus. Ei ole vain yhtä tekoölyä, vaan monia sovelluksia, järjestelmiä, menetelmiä, ohjelmia ja työkaluja, joissa on mukana tekoölyä. Tekoölysovellukset perustuvat erilaisiin teknisiin ratkaisuihin ja niitä kehitetään eri tarkoituksiin. Käsitteet tekoölystä muuttuvat myös ajan

¹² Tekoölyä käsittelevä korkean tason asiantuntijaryhmä, 2019

¹⁴ Ruckenstein, 2023, s. 3

¹⁵ Ruckenstein, 2023, s. 13

¹³ Gillespie, 2023

¹⁶ UNESCO, OECD, IDB, 2022

kuluessa. Esimerkiksi tekoäly tutkimusaiheena määrittyy jatkuvasti uudelleen, kun uusia sovelluksia syntyy ja joidenkin aihepiirien ei katsota enää kuuluvan sen piiriin. Yksi esimerkki on syväväärentäminen. Syväväärennös, eli niin kutsuttu deepfake, on aidolta vaikuttavaa kuvaa tai ääntä sisältävä kuva- tai videoväärennös, jota kutsuttiin aiemmin kuvankäsittelyksi mutta nykyään tekoälyksi¹⁷. Syväväärennösten avulla ihminen voidaan saada sanomaan ja tekemään asioita, joita hän ei ole todellisuudessa tehnyt tai sanonut.

Tässä selvityksessä tekoäly ymmärretään Euroopan neuvoston tekoälykomitean (Ad hoc Committee on Artificial Intelligence, CAHAI) määritelmän mukaisesti kattoterminä erilaisille tietokoneella käytettäville sovelluksille, jotka muistuttavat toiminnoiltaan ihmisen älykkyyttä¹⁸. Hannu Toivosen mukaan tekoälyn älykkyydessä on kuitenkin kyse enemmän kielikuvasta, jolla tietokoneohjelman toiminta rinnastetaan inhimilliseen älykkäiseen toimintaan. Kyse ei ole niinkään älykkyydestä kuin jonkinlaisesta vertailukelpoisuudesta älykkyyden kanssa¹⁹. Määritelmässä voidaan korostaa myös prosessia: tekoäly näyttäytyy ihmisen luomina järjestelminä, joiden avulla pyritään asetettuihin tavoitteisiin keräämällä, tulkitsemalla ja prosessoimalla digitaalisia aineistoja. Pyrkimyksenä on tehostaa itseohjautuvuutta, organisointia ja yhteiskunnan toimintoja²⁰.

Koneoppiminen

Koneoppiminen on eräs tekoälyn soveltamisen osa-alue, joka perustuu säännönmukaisuuksien oppimiseen datasta. Koneoppimisessa tietokoneelle syötetään dataa, josta se oppii ja pystyy tekemään itsenäisesti erilaisia päätelmiä, kuten ennusteita. Koneoppimisessa tietokone siis etsii yhteneväisyyksiä ja korrelaatioita suurista tietojoukoista eli luokittelee dataa. Koneoppimista voi toteuttaa joko niin, että ihmiset ovat määritelleet luokat etukäteen tietokoneelle tai että ne syntyvät vasta koneoppimisen tuloksena.

Eräs koneoppimiseen yhdistetty haaste on niin kutsuttu palautekehäongelma (feedback loop problem). Se viittaa tilanteeseen, jossa tekoälyjärjestelmän tekemät päätökset tai ennusteet vaikuttavat tulevaan dataan, jota järjestelmä käyttää oppiakseen ja parantaakseen itseään. Tämä voi johtaa itseään vahvistavaan sykliin, jossa alkuperäiset ennusteet ja päätökset vaikuttavat tulevaan käyttäytymiseen ja dataan, ja uusi data puolestaan vahvistaa alkuperäisiä ennusteita ja päätöksiä. Yhdysvalloissa tekoälyä on sovellettu tällä tavalla rikollisuuden torjuntaan²¹. Esimerkiksi jos tekoälyjärjestelmä tunnistaa jollakin alueella paljon rikollisuutta, sinne saatetaan ohjata lisää poliisivoimia. Lisääntyneen poliisivoiman seurauksena rikokset todennäköisesti myös havaitaan herkemmin kuin alueella, jolla ei ole poliiseja. Näin ollen myös kirjattavien rikosten määrä kasvaa tällä alueella ja ennuste vahvistuu.

Algoritmit

Algoritmi on yksityiskohtainen kuvaus tai ohje, jota seuraamalla tekoälyjärjestelmä suorittaa tehtävän, prosessin tai ratkaisee ongelman. Esimerkkinä voidaan käyttää yllä kerrottua esimerkkiä siitä, miten koneoppimisessa tietokoneelle voidaan määritellä

¹⁷ Lanier, 2020

²⁰ Ruckenstein, 2022a, s. 3

¹⁸ Bartoletti & Xenides, 2023, s. 13 ²¹ D'Ignazio & Klein, 2023

¹⁹ Toivonen, 2023

etukäteen luokat, joihin dataa luokitellaan. Yleisellä tasolla algoritmit voi nähdä monimutkaisina reseptikimppuina, jotka sisältävät ohjeita tehtävien suorittamiseen ja ongelmien ratkaisuun²². Arkisessa kielenkäytössä ehkä tutuin tapa viitata algoritmeihin liittyy suosittelujärjestelmiin (recommender systems), joiden perusteella saamme räätälöityjä suosituksia sosiaalisen median alustoilla ja elokuva- ja musiikkisuosituksia suoratoistopalveluissa, kuten Netflixissä ja Spotifyssa²³.

Kielimallit ja generatiivinen tekoäly

Kielimallit perustuvat sanojen ja kirjoitusmerkkien tilastollisiin yhteyksiin. Ne tekevät niille annetun aineiston perusteella ennustuksia, jotka toistavat tilastollisia säännönmukaisuuksia. Suurilla aineistoilla koulutetuilla kielimalleilla voi tuottaa eli generoida tekstiä. Kielimallisovelluksia nimitetäänkin myös generatiiviseksi tekoälyksi. Generatiivinen tekoäly viittaa tekoälyjärjestelmiin, jotka pystyvät tuottamaan uutta, alkuperäistä sisältöä. Tämä sisältö voi olla tekstiä, kuvia, musiikkia, koodia tai muita tietomuotoja. Generatiivinen tekoäly oppii tuottamaan tällaista sisältöä analysoimalla suuria määriä olemassa olevaa dataa ja oppimalla sen rakenteista ja piirteistä.

Kielimallit tuottavat yleensä sujuvan oloista tekstiä, vaikka sovellus itse ei ymmärrä käyttämiensä sanojen merkitystä muuten kuin yhteyksinä toisiinsa. Viime aikoina kielimallit ovat dominoineet keskustelua tekoälystä jopa siihen pisteeseen, että monet pitävät niitä samana asiana. OpenAI:n ChatGPT:n eri versiot nostivat kielimalleihin perustuvan tekoälyn uutisoinnin keskipisteeseen ja viime aikoina teknologiajätit ovat kehittäneet niille kilpailijoita, joille mietitään kuumeisesti eri käyttötarkoituksia.

Kaikkiin yllä oleviin määritelmiin liittyy epävarmuutta ja hämmennystä. Selvitystä taustoittavissa keskusteluissa korostettiin, että eri tieteenaloilla, mutta myös arkikäytössä, sanat kuten tekoäly, koneoppiminen ja algoritmit tarkoittavat eri asioita. Määritelmistä päättäminen on myös vallankäyttöä: niihin liittyy aina painotuksia, joiden avulla voi edistää päämääriä. Selvityksen tavoitteena ei ole löytää konsensusta määritelmille eikä niiden painotuksille. Jaetut määritelmät ovat tarpeellisia tarkkaan rajatussa asiantuntijatyössä, mutta laaja-alaisissa yhteiskunnallisissa keskusteluissa on hyvä säilyttää joustavuus ja pyrkiä monitieteiseen lähestymistapaan.

Vinoutunut malli ongelmanratkaisulle. Tekoälyä on kehitetty jo vuosikymmeniä ja esimerkiksi ajatukset siitä, miten teknologia saadaan mallintamaan ja toimimaan kuten ihmisten ongelmanratkaisu tai äly, perustuvat paljolti 1970-luvulla tehtyihin psykologisiin kokeisiin. Näissä kokeissa tavoiteltiin tietoa ihmisten yleisistä ongelmanratkaisu- ja kyvyistä. Ongelmallisen asetelmasta teki se, että kokeisiin osallistui lähinnä teknisen yliopiston opiskelijoita Yhdysvalloissa. Tuohon aikaan he olivat pääosin hyvätulaisia, matemaattisesti lahjakkaita, valkoisia, amerikkalaisia, nuoria miehiä. Tällöin yleisen ongelmanratkaisun esimerkkinä käytettiin hyvin rajattua ihmisryhmää²⁴.

²² Ruckenstein, 2023, s. 3

²⁴ Adam, 1998

²³ Bucher, 2016

2. Tekoälyn toisintamat sukupuolistereotyytiat ja syrjintä

Tässä luvussa lähestytään tekoälyn syrjiviä vaikutuksia suhteessa sukupuolten tasa-arvoon erilaisten tapausten ja esimerkkien kautta. Luvun alussa perehdytään siihen, millaisia sukupuolta koskevan datan vinoumia tekoäly tuottaa, miten sukupuolta luokitellaan ja kuinka data ei koskaan ole neutraalia. Tekoälyjärjestelmät perustuvat pääosin datan luokitteluun ja analysointiin, minkä takia on syytä tarkastella datassa esiintyviä vinoumia ja puutteita. Datan tarkastelun kautta voimme ymmärtää ja arvioida tekoälysovellusten tekemiä päätelmiä ja niiden käytettävyyttä. Tämän jälkeen luvussa käsitellään sukupuoleen liittyvää syrjintää arjessa, kuten sosiaalisen median algoritmeissa, ja miten sukupuoli näkyy kielimalleissa ja toimii tekoälyjärjestelmien ennusteissa. Lopuksi keskustelua laajennetaan kohti laajoja rakenteellisia muutoksia ja dataetiikkaa.

2.1 Datan vinoumat, puutteet ja luokittelu

Kansainvälisessä kirjallisuudessa käytetään termiä gender data gap kuvaamaan monilla yhteiskunnan sektoreilla ilmenevää rakenteellista vinoumaa: dataa on kerätty tai sitä on saatavilla lähinnä miehistä. Esimerkiksi lääketieteellistä tutkimusta ja erityisesti klinisiä lääketutkimuksia on tehty perinteisesti miesten parissa. Naiset eivät ole olleet klinisten tutkimusten kohteena useasta syystä²⁵. Ensinnäkin keskittymällä miehiin on pyritty välttämään, etteivät lisääntymisiässä olevat naiset osallistuisi mahdollisesti sikiölle haitallisiin tutkimuksiin. Toisaalta lääketieteessä on vielä viime vuosisadalla ollut vallalla ajattelumalli, jossa miesruumis on eräänlainen standardi-ihminen. Ajateltiin, että erot naisten ja miesten terveydessä ja sairastamisessa liittyivät lisääntymisterveyteen, eikä naisten oireita tai taudin kuvaa tarvitse tutkia erikseen.

Systemaattiset vääristymät ja puutteet kliinisessä datassa voivat vaikuttaa siihen, millaisia yhteyksiä tekoäly tunnistaa oireissa tai millaisia ennusteita sairauden etenemiselle se tekee. Lääketieteellisen tutkimuksen sukupuolivinouman takia tämä voi erityisesti vaikuttaa naisiin ja heidän hoitoonsa, mutta myös muihin sellaisiin ihmisryhmiin, jotka ovat olleet aliedustettuina lääketieteellisissä tutkimuksissa.

Tätä vinoumaa on pyritty korjaamaan uusilla säännöksillä, jotka velvoittavat erilaisten ihmisryhmien mukaan ottamista lääketieteelliseen tutkimukseen. Rakenteellista vinoumaa ei kuitenkaan voida poistaa nopeasti. Vaikka uusiin tutkimuksiin otettaisiinkin enemmän naisia ja vähemmistöryhmien edustajia mukaan, lääketiede tarvitsee vertailudataa ja

²⁵ Lego, 2023

aikaisempi data on kerätty lähinnä miehistä. On tärkeä ymmärtää, että vinoumat datassa ja algoritmeissa eivät aina tarkoita syrjintää. Mutta vinoumaherkkyys on tekijä, joka luo tekniset edellytykset algoritmin aidosti syrjivälle käytölle²⁶.

Lääketiede toimii esimerkkinä sektorista, jossa rakenteelliseen vinoumaan ja datan puutteeseen on havahduttu. Kaikkia puutteita ja vinoumia ei kuitenkaan aina tunnisteta. David Hand²⁷ käyttää termiä pimeä data kuvaamaan dataa, jota ei ole tai johon ei ole pääsyä, mutta joka vaikuttaa tai jonka tulisi vaikuttaa päätelmiin, joita aineistosta tehdään. Pimeä data on tietoa, jonka tiedämme puuttuvan, mutta myös dataa, jonka puuttumisesta emme tiedä. Pimeällä datalla voidaan siis kuvata esimerkiksi aineistoa, joka sisältää vääristymiä, puolueellisuutta tai puuttuvaa tietoa ilman, että käyttäjät ovat siitä tietoisia. Joskus puuttuvan datan tunnistaa vasta vuosikymmenien päästä. Tämä on tärkeä havainto, koska ennusteet, joita tekoälyllä tehdään tulevaisuudesta perustuvat nykyhetkessä tai aikaisemmin kerättyyn dataan.

Luokittelut väestötilastoissa. Merja Kinnunen on tutkinut kuinka työhön sekä taloudelliseen ja yhteiskunnalliseen asemaan liittyvät luokittelut ovat ylläpitäneet Suomessa sukupuolten välistä jakoa²⁸. Kinnunen nostaa esiin kuinka yhtä pitkän koulutuksen vaativat naisenemmistöiset ammatit (kuten sairaanhoitajat) luokiteltiin useammin alempien toimihenkilöiden luokkaan ja miesenemmistöiset ammatit (kuten insinöörit) ylempiin toimiluokkiin. Pitkälle 1900-luvun puolivälin jälkeen väestötilastoissa käytettiin myös erilaisia avustavan perheenjäsenen luokkia kuvaamaan yleensä naisia, jotka toimivat maataloilla tai yrityksissä, jolloin heidän asemansa määrittyi miehen aseman mukaan. Vaikka kaikki sukupuolittavat kategoriat eivät ole enää käytössä, vanhojen kategorioiden ymmärtäminen on tärkeää, koska kuten Kinnunen toteaa: ”nykyisyyden ymmärtämisessä ja tulevaisuuden arvioinnissa ovat läsnä käsitykset yhteiskunnan menneisyydestä²⁹.”

Tarpeellista ja luotettavaa sukupuolen mukaan eriteltyä dataa ei aina siis ole saatavilla. Tämä liittyy myös kysymykseen siitä, miten sukupuoli luokitellaan. Tietojenkäsittely itsessään perustuu luokitteluun, eikä dataa voi olla ilman, että sitä on luokiteltu jollain tavalla³⁰. Samoin kaikki nykyaikaiset organisaatiot toimivat tiedon ja tietojärjestelmien avulla, ja kaiken perustana on tiedon luokittelu. Luokittelu on välttämättömyys tiedon käsittelyn ja arkisen toiminnan kannalta. Tutut käytännöt ja luokittelut puheessa sujuvoittavat keskustelua: puhumme vihreistä banaaneista, mutta emme erikseen keltaisista, koska oletuksena on, että banaanit ovat keltaisia. Samoin kielemme on sukupuolittunut. Puhumme esimerkiksi miehistä ja naisista, vaikka tiedostamme sukupuolen moninaisuuden. Luokittelussa on kyse valinnoista – mitä luokitellaan, mitä kategorioita, ominaisuuksia ja luokitteluperusteita käytetään.

²⁶ Rusanen, 2021

²⁹ Kinnunen, 2022, s. 69

²⁷ Hand, 2020

³⁰ D'Ignazio & Klein, 2023, s. 103

²⁸ Kinnunen, 2022

Sukupuolen mukaan eritelty data, jota on julkisissa rekistereissä sekä monissa yksityisissä palveluissa, on yleensä binääristä eli luokitteluvaihtoehtoina on vain mies ja nainen. Esimerkiksi Suomen väestötietojärjestelmä ei tarjoa muita vaihtoehtoja. Binäärisen luokittelun takia niin Suomessa kuin kansainvälisesti on saatavilla vain vähän dataa sukupuolivähemmistöistä siitäkkin huolimatta, että joissain maissa on virallisesti käytössä kolmas juridinen sukupuoli. Viime aikoina esiin on kuitenkin noussut tarve saada tietoja eri vähemmistöryhmistä, kuten sukupuolivähemmistöistä³¹. Tällaisen tilastotiedon avulla voitaisiin pyrkiä tunnistamaan ja korjaamaan yhteiskunnassa ilmeneviä syrjiviä rakenteita.

Joskus tieto sukupuolesta ja erityisesti sukupuolivähemmistöön kuulumisesta voi johtaa syrjintään. On tilanteita, joissa voi olla parempi jättää sukupuoli kokonaan ilmoittamatta. Lisäksi joissain tapauksissa voi olla parasta jättää sukupuolen tarkemmat määrittelyt (kuten cis- ja trans-luokittelut) tekemättä. Sukupuoliluokittelut ovat kuitenkin usein pakollisia sovelluksiin kirjautuessa, sillä ne määrittävät osaltaan suositusjärjestelmien toimintaa ja sisältävät muita niihin ohjelmoituja taustaoletuksia. Tällöin sukupuolille suositellaan tai esitetään erilaisia sisältöjä.

Ei siis ole vain yhtä oikeaa tapaa suhtautua luokitteluun sukupuolten tasa-arvoa edistävällä tavalla, vaan sopivin tapa pitää määritellä tilannekohtaisesti. Kyseessä on aina tasapainottelu etujen ja haittojen välillä kulloisessakin yhteiskunnallisessa kontekstissa. Tarkkojen yksilöä määrittävien tietojen kerääminen ja käyttö tekoälyn aineistoissa voi asettaa henkilöitä vaaraan, jos heidät voi tunnistaa niiden perusteella. Samalla kuitenkin huolellisesti suunniteltu ja vähemmistöihin kuuluvien henkilöiden suostumukseen perustuva luokittelu voi edistää heidän asemaansa tuomalla ryhmän esiin ja lisäämällä sen näkyvyyttä. Puuttuvan ja arkaluonteisen datan ongelmaan on ehdotettu ratkaisuksi synteettisen datan käyttöä³². Synteettisellä datalla tarkoitetaan keinotekoisesti luotua dataa – eli vaihtoehtoa todelliselle datalle. Synteettinen data voidaan tehdä vastaamaan tilastollisesti alkuperäistä aineistoa, mutta niin että yksilötasolla data ei kuitenkaan vastaa yksittäisen henkilön tietoja.

Erilaiset lasku- ja luokittelutavat eivät automaattisesti johda syrjintään – niillä voi myös edistää solidaarisuutta, tuoda esiin pimennossa olleita asioita historiasta ja haastaa vallitsevia rakenteita³³. Aivan kuten yllä käsitellyt data-aineistot, tavat laskea ja luokitella eivät ole koskaan neutraaleja.

³¹ Salokannel & Ignatius, 2023

³² Musizvingoza, 2024

³³ D'Ignazio & Klein, 2023, s. 118

Tekoäly ja rikollisuus. Vuonna 2019 Lontoon poliisi suoritti kokeilun, jossa käytettiin reaaliaikaista kasvontunnistussovellusta tunnistamaan väkivaltarikoksista tuomittuja ja terrori-iskujen suunnittelusta epäiltyjä henkilöitä kadulla. Kokeilun aikana havaittiin, että järjestelmä oli altis vinoumille ja syrjinnälle. Sovellus oli taipuvainen luokittelemaan ei-valkoihoisia henkilöitä rikollisiksi, silloinkin kun he eivät sitä olleet. Valkoihoisten henkilöiden kohdalla vastaavaa virhepäätelmää ei tapahtunut ja sovellus toimi tarkemmin. Tämä tarkoitti, että järjestelmän käyttö johti syrjiviin tilanteisiin, joissa ihmisiä epäiltiin rikollisiksi pelkästään ihonvärin perusteella³⁴.

2.2 Miksi data ei ole koskaan neutraalia?

Arkikielessä puhutaan usein raakadatasta, millä tarkoitetaan dataa, joka mielletään neutraaliksi ja koskemattomaksi. Ruha Benjamin viittaa kirjassaan *Race after Technology: Abolitionist Tools for the New Jim Code*³⁵ ilmiöön, jota hän kutsuu datan ja teknologian kuvitteelliseksi objektiivisuudeksi. Se tarkoittaa, että määrällinen data käsitetään objektiivisena, koska se ei ole altista ihmisten tulkinnoista nouseville vääristymille. Samaa ilmiötä voidaan kutsua myös hieman eri painotuksilla automaatiivääristymäksi (automation bias)³⁶ tai algoritmivääristymäksi (algorithm bias)³⁷.

Samanlaista puhetta objektiivisuudesta esiintyy myös tekoälyn kohdalla. Data – ja silloin myöskään tekoäly – ei kuitenkaan ole koskaan puhdasta, raakaa tai objektiivista vaan sen kerääminen ja analysoiminen on aina inhimillisten valintojen ja luokittelujen – toisin sanoen datatyön – tulosta. Data ei synny itsestään vaan ensin pitää valita minkälaista dataa kerätään, miten sitä mitataan, minkälaista mittayksikköä tai luokittelua käytetään ja mitä jätetään ulkopuolelle. Tämän jälkeen data vaatii yleensä käsittelyä, kuten aineiston yhdenmukaistamista, päällekkäisyyksien poistamista tai puuttuvien tietojen lisäämistä tai päättelemistä. Rekisteriaineistot eivät esimerkiksi usein salli ihmiselle kahta ammattia tai äidinkieltä, jolloin ihminen tai algoritmi valitsee näistä kirjattavaksi vain toisen. Kun aineistoa korjataan, voidaan puhua imputoinnista. Tällöin rekisterin sisältämää puuttuvaa tietoa paikataan päättelyllä joko laskennallisin tai inhimillisen päättelyn keinoin³⁸.

³⁴ Rusanen, 2021

³⁷ Garcia, 2016

³⁵ Benjamin, 2019

³⁸ Alastalo, 2009; Alastalo & Homanen, 2015

³⁶ Bartoletti & Xenides, 2023, s. 18

Tekoäly ja rationaalisuus. Data ja tekoäly yhdistetään puheessa usein objektiivisuuteen ja rationaalisuuteen. Samanaikaisesti rationaalisten ominaisuuksien liittäminen miehisyyteen ja tunteellisten naisellisuuteen on yksi monista syvään juurtuneista stereotyypeistä, joka vaikuttaa puheeseen teknologiasta ja sen soveltamisesta yhteiskunnassa. Uskomuksen pysyvyydestä kertoo sekin, että vaikka monet voimakkaatkin tunteet, kuten viha ja ylpeys, liitetään yleisesti miehiin, tunteet yleisellä tasolla yhdistetään naisiin³⁹. Tämä paikkaansa pitämätön jaottelu ohjaa myös yleisesti jaettuun käsityksiä datasta ja objektiivisuudesta. Sukupuolittava stereotyyppinen näkemys siis samaistaa yhteismitallisen määrällisen datan miehiseen objektiivisuuteen ja laadullisen sosiokulttuurisen tulkinnan tunteelliseen naisellisuuteen.

2.3 Algoritminen syrjintä ja stereotyyppien ylläpito

Monesti tekoälyn syrjivät vaikutukset jäävät huomaamatta, koska erilaisista tekoälyä käyttävistä sovelluksista on tullut osa ihmisten ja organisaatioiden arkea. Esimerkiksi ruokakaupan hintataso ja valikoima olisivat toisenlaisia ilman sitä – tekoälyä käyttävät ohjelmistot ovat nykyään läsnä niin maataloudessa, ruuan tuotannossa, sen kuljettamisessa varastoille ja kauppoihin sekä itse myymälöissä⁴⁰. Kaikenlaisen datan kerääminen kansalaisista julkisissa ja yksityisissä palveluissa, sosiaalisessa mediassa ja digitaalisten laitteiden avulla on yhä halvempaa, helpompaa ja samalla tiiviimmin kietoutunutta arkiseen elämään. Tekoälyn kehittäjien näkökulmasta ihmiset edustavat potentiaalisesti louhittavaa dataa, jonka avulla heidän reaaliaikaista käyttäytymistään voidaan seurata ja jopa valvoa. Kerättyä dataa voidaan käyttää poliittisiin ja taloudellisiin päämääriin uusilla, huomaamattomilla tavoilla⁴¹.

Myös sosiaalisen median alustojen ja sovellusten käyttämä tekoäly on vaikeasti paikannettavissa. Silti algoritmit ja suosittelujärjestelmät ovat osa jokaista Google-hakua ja verkossa eteen osuvaa mainosta⁴². Näkymättömyyden takia algoritmien vaikutukset sukupuolten tasa-arvoon ovat kuitenkin voimakkaita juuri sosiaalisessa mediassa. Sosiaalisen median, kuten Instagramin, TikTokin ja Facebookin toimintalogiikka perustuu algoritmeihin, joiden avulla pyritään tuottamaan jokaiselle käyttäjälle henkilökohtaista sisältöä. Algoritmit on suunniteltu pitämään käyttäjät sitoutuneina alustaan. Yksi tunnistettu ongelma liittyy kehonkuvaan. Tutkimusten mukaan⁴³ mukaan sosiaalisen median käytön ja vääristyneen kehonkuvan välillä on yhteys. Sosiaalisen median algoritmit voivat voimistaa tätä yhteyttä, tarjoten samankaltaista, mutta usein äärimmäisempää sisältöä, kuin mitä käyttäjä on jo katsonut tai lukenut.

Vaikka sosiaalisen median alustat antavat käyttäjilleen mahdollisuuden vaikuttaa siihen, millaista sisältöä heille näytetään, esimerkiksi valitsemalla seurattavat henkilöt ja tilit,

³⁹ D'Ignazio & Klein, 2023, s. 77

³⁹ D'Ignazio & Klein, 2023, s. 77

⁴⁰ Toivonen, 2023, s. 22

⁴¹ Ruckenstein, 2023, s. 6

⁴² Larsson et al., 2024

⁴³ Kts esim. Harriger et al., 2022

käyttäjät ovat kuitenkin usein tietämättään sosiaalisen median algoritmien vaikutuksen alaisia. Silloin jopa käyttäjät, jotka pyrkivät välttämään tietynlaista sisältöä tai seuraamaan positiivisempia tilejä, eivät aina ole suojassa. Algoritmien avulla voi siis helposti levitä vääristyneitä kuvia esimerkiksi siitä, millaisia naisten ja miesten kehojen tulisi olla. Tämä voi puolestaan voimistaa sukupuoleen liittyviä stereotyyppioita ja normatiivisia odotuksia.

Erityisesti sosiaalisen median ja hakukoneiden algoritmit voivat luoda niin sanotun kuplautumisen ilmiön, jossa käyttäjät altistuvat enimmäkseen sisällölle, joka vahvistaa heidän olemassa olevia uskomuksiaan ja mielipiteitään. Tämä voi vahvistaa käyttäjien ennakkoluuloja ja vihamielisiä asenteita. Myös suomalaisessa mediassa⁴⁴ on nostettu esiin ilmiö, jossa nuoret miehet altistuvat naisvihamieliselle sisällölle. Samassa yhteydessä puhutaan ”kaikukammioista” eli ilmiöstä, jossa samat ajatukset ja näkemykset kaikuvat takaisin käyttäjille ilman vastakkaisia näkemyksiä⁴⁵.

Vuonna 2018 teknologiajätti Googlen avoimessa AI Experiments –hankkeessa kehitettiin piirtämiseen perustuva Quick, Draw! –peli, jossa osallistujat piirsivät arkisia esineitä, jotka tekoäly sitten tunnisti ja luokitteli⁴⁶. Aluksi osallistujat piirsivät lenkkareita kuvaamaan tyypillisiä kenkiä, ja tekoälymallia koulutettiin näiden kuvien perusteella. Tämän seurauksena valmis peli ei tunnistanut esimerkiksi naisten korkokenkiä kengiksi. Tämä esimerkki saattaa vaikuttaa vähäpätöiseltä, mutta se kuvaa hyvin toimintalogiikkaa tekoälyn ja algoritmien taustalla. Pienetkin vinoumat tekoälyjärjestelmissä voivat kasvaessaan muodostaa merkittäviä esteitä tasa-arvon toteutumiselle. Jos sukupuoleen liittyvät vääristymät jäävät huomaamatta ja korjaamatta, ne voivat heijastua laajemmin yhteiskunnassa. Näin ollen on ensiarvoisen tärkeää, että tekoälyjärjestelmien kehitystyössä otetaan huomioon moninaiset näkökulmat ja pyritään aktiivisesti poistamaan vinoumat, jotta voidaan edistää sukupuolten tasa-arvoa ja oikeudenmukaisuutta.

Tekoälyyn ja algoritmeihin liitetään myös usein käsite profilointi. Sillä tarkoitetaan prosessia, jossa algoritmit analysoivat suuria määriä dataa eri ihmisistä ja luovat näiden tietojen perusteella profiileja. Eli ihmisiä luokitellaan luokkiin käyttäytymisen, mieltymysten ja muiden piirteiden perusteella. Eräs tunnettu esimerkki, jossa algoritmit profiloivat sukupuolen perusteella, on Googlen mainostyökalujen käyttö rekrytointi-ilmoituksissa. Tutkijat havaitsivat⁴⁷, että Googlen mainosalgoritmit näyttivät korkeapalkkaisia työpaikka-ilmoituksia useammin miehille kuin naisille. Tutkimuksessa käytettiin keinotekoisia käyttäjäprofiileja, joiden avulla tutkijat simuloivat verkkokäyttäytymistä ja analysoivat, millaisia mainoksia eri profiileille näytettiin. Tämä sukupuoleen pohjautunut profilointi johtui osittain siitä, että mainosalgoritmit optimoivat mainosten näyttämistä aiemman verkkohistorian ja muiden käyttäjätoimintojen perusteella. Koska miehet olivat useammin päätyneet korkeapalkkaisten työpaikkojen mainossivuille, algoritmi oppi näyttämään näitä mainoksia useammin miehille. Tämä johti siihen, että naiset saivat vähemmän mahdollisuuksia nähdä ja hakea korkeapalkkaisia työpaikkoja, mikä vahvisti olemassa olevia sukupuolten välisiä eroja työmarkkinoilla. Tapaus herätti laajaa huomiota ja keskustelua siitä, kuinka mainosalgoritmit voivat vahvistaa ja ylläpitää sukupuolittuneita eroja ja syrjintää.

⁴⁴ Similä, 2024

⁴⁷ Datta et.al, 2015

⁴⁵ Mogensen & Holding Rand, 2020

⁴⁶ Lovejoy, 2018

Tilastollisten menetelmien tuottama syrjintä luottoluokituksissa

Yhdenvertaisuusvaltuutetun pyynnöstä yhdenvertaisuus- ja tasa-arvolautakunta käsitteli tapausta, joka koski automaattista päätöksentekoa luotonannossa⁴⁸. Asianosainen mieshenkilö oli hakenut luottoa hankkimiansa talotarvikkeiden rahoittamiseksi, mutta saanut kielteisen luottopäätöksen. Päätöksen perusteluja koskevaan tiedusteluun miehelle oli vastattu, että päätös oli perustunut ”luottotietopalveluiden tilastollisiin menetelmiin perustuviin luottoluokituksiin, jotka eivät ota kantaa yksittäisen luotonhakijan maksuvalmiuteen, ja saattavat poiketa paljonkin luotonhakijan profiilista sekä tuntua luotonhakijasta epäoikeudenmukaisilta⁴⁹.”

Luotonhakijalla ei ollut maksu- tai luottohäiriömerkintöjä eikä luottohakemuslomakkeessa kysytty hänen tulo- ja varallisuustietojaan. Sitä vastoin pisteytysjärjestelmä perustui muuan muassa luotonhakijan asuinpaikkaan, sukupuoleen, äidinkieleen ja ikään. Kuten monissa muissa selvityksen esimerkeissä, tekoälyjärjestelmien tuottama syrjintä ei perustunut yksinomaan sukupuoleen, vaan useisiin risteäviin tekijöihin, jolloin kyse oli niin kutsutusta moniperusteisesta syrjinnästä. Tässä tapauksessa sukupuoli vaikutti seuraavalla tavalla: miehillä oli järjestelmässä enemmän maksuhäiriöitä kuin naisilla, joten he saivat vähemmän pisteitä kuin naispuoliset luotonhakijat. Järjestelmässä myös suomenkieliset hakijat saivat vähemmän pisteitä kuin ruotsinkieliset, ja ruotsinkielinen nainen olisi saanut hakijan muilla tiedoilla myönteisen luottopäätöksen. Järjestelmä ei siis tarkastellut yksittäisen hakijan todellista tilannetta, vaan teki tilastollisen arvion saamiensa tietojen pohjalta.

Yhdenvertaisuusvaltuutetun saaman selvityksen perusteella syntyi epäily, että hakijan yksilöllistä luottokelpoisuutta ei ollut tarkastettu. Valtuutettu nosti esiin EU-tuomioistuinten aiemman ratkaisun, jossa kiellettiin sukupuolen käyttäminen vakuutusmatemaattisena muuttujana. Tähän luotonantaja vastasi viittaamalla lainsäätäjän edellyttämään luotonantajan arvioon kuluttajaluotonhakijoiden maksukyvyistä ja esitti, että ”lainsäätäjä on nimenomaan hyväksynyt tilastollisten menetelmien käytön osana luotonmyöntöä, sillä se on menetelmä, joka kohtelee kuluttajia tasapuolisesti⁵⁰”.

Vuonna 2018 asiasta tehtiin päätös, jossa yhdenvertaisuus- ja tasa-arvolautakunta kielsi luotonantajaa uusimasta luotonhakijaan tai kehenkään muuhun päätöksestä ilmenevää yhdenvertaisuuslain vastaista menettelyä. Päätöksessä katsottiin, että luotonantaja ei ollut kyennyt kumoamaan syntyneitä syrjintäolettamaa, vaan toteutti moniperusteista syrjintää hakijan sukupuoleen, äidinkieleen, ikään ja asuinpaikkaan liittyvien syiden perusteella.

⁴⁸ Ruckenstein, 2022b

⁴⁹ Yhdenvertaisuus- ja tasa-arvolautakunta, 2018

⁵⁰ Yhdenvertaisuus- ja tasa-arvolautakunta, 2018, s. 12

Sukupuolittuneet virtuaaliset apurit. Tekoälyn muodot voivat myös ylläpitää sukupuolittuneita asenteita, koska niissä yhdistetään feminiinisinä pidettyjä ominaisuuksia tiettyihin tehtäviin. Tutkimuksissa on esitetty⁵¹, että naiseksi nimetyt tai naisellista ääntä käyttävät tekoälysovellukset koetaan inhimillisemmiksi ja helpommin lähestyttäviksi. On myös esitetty, että naisäänellä varustetulle tekoälylle jaetaan helpommin sensitiivistäkin tietoa kuin miesäänelle. Tämä on huomioitu sovellusten kaupallisessa kehittämisessä. Esimerkiksi suurten yritysten kehittämät ääniohjautuvat apurit on nimetty Alexaksi, Siriksi ja Cortanaksi. Miesten nimet ja äänet loistavat poissaolollaan. Ivana Bartoletti esittää, että tämä on tapa normalisoida sukupuolirooleja: aiemmat piit, sisäköt ja sihteerit ovat saaneet näistä uudet digitaaliset vastineensa, jotka toteuttavat kuuliaisesti ja mukisematta niille annettuja tehtäviä⁵². Valinnat vahvistavat sukupuolistereotypioita, joissa naiset ovat tottelevaisia, avuliaita ja miellyttäviä. Joissakin tapauksissa apurit on jopa ohjelmoitu niin, että ne vaikuttavat olevan imarreltuja seksuaalisesta häirinnästä tai hyökkäävästä kohtelusta⁵³.

2.4 Generatiivinen tekoäly ja syrjintä

Tekoälyjärjestelmä toimii sille annettujen aineistojen ja kommentojen perusteella. Generatiivinen tekoäly käyttää usein hyödykseen laajoja, suodattamattomia aineistoja, joita käytettäessä toistetaan helposti sukupuolistereotypioita. Eli jos taustalla oleva teksti- tai kuva-aineisto sisältää puhetta miehistä johtajina ja naisista perheenäiteinä, tekoäly generoi samankaltaisia tuloksia. Samoin kuvahaku sanalla ”lääkäri” nostaa esiin lähinnä miespuolisia lääkäreitä – naislääkäri on ollut historiallisesti poikkeus normaalista, joten hakukone ei huomioi sitä, ellei sitä erityisesti ohjelmoida tekemään niin. Google Translate -kääntäjä taas on saanut kritiikkiä siitä, että se on käyttänyt sukupuolittuneita käännoiksiä sukupuolineutraaleille sanoille. Esimerkiksi suomenkielinen lause ”hän on lääkäri” kääntyi aikaisemmin englanniksi vain muotoon ”he is a doctor”. Nykyään käännohjelma antaa käännoksiin kaksi vaihtoehtoa she ja he -pronomineilla. Yhtä lailla, tekoälyn luoma kuva kengistä rajoittuu miesten lenkkareihin, jos sitä ei kouluteta monipuolisemmilla aineistoilla. Tekoäly tekee rajallisia ja usein syrjiviä ennustuksia yksipuolisen ja menneisyyttä edustavan datan perusteella⁵⁴. Se ei pysähdy pohtimaan erikseen stereotypioista poikkeavia tuloksia.

Hyvä esimerkki huomaamattomasta epätasa-arvoa ylläpitävästä aineistosta on Wikipedia, joka on merkittävä lähde monien kielimallien, kuten ChatGPT:n eri versioiden, koulutuksessa. Miehet ovat olleet yliedustettuina sekä Wikipedian kehittäjinä että tuotetussa sisällössä. Wikipedian puolestatoista miljoonasta tunnettuja henkilöitä esittelevästä

⁵¹ Morsello & Bonono, 2024; Woods, 2018

⁵⁴ D'Ignazio & Klein, 2023

⁵² Bartoletti, 2020

⁵³ EIGE, 2021

sivusta vain alle viidennes käsittelee naisia. Tämä epätasapaino vaikuttaa merkittävästi kielimallien tuottamiin historiallisiin ja sosiokulttuurisiin tiivistelmiin eri aiheista.

Monet yritykset ovat ryhtyneet poistamaan vinoumia aineistostaan vahvistamalle esimerkiksi naisten näkyvyyttä aineistossa. Tämän seurauksena on itseasiassa ajauduttu uudenlaisiin vinoumiin⁵⁵. Korjaustyössä perinteisesti maskuliinisia ominaisuuksia liitettiin myös naisiin, mutta feminiinisiä ominaisuuksia ei vahvistettu miesten kohdalla. Tämä malli on linjassa tutkimuksien kanssa, jotka osoittavat sosiaalista haluttomuutta hyväksyä poikia, jotka osallistuvat perinteisesti naiselliseen toimintaan, ja vastareaktion niitä miehiä vastaan, jotka uhmaavat sukupuolinormeja.

Generatiivisen tekoälyn pohjana käytetyt tietoaaineistot esimerkiksi sosiaalisen median alustoilta, kuten X:stä tai Redditistä, ohjaavat kielimallien tuottamia vastauksia ja kuvauksia syrjiviin ja usein loukkaaviin suuntiin, ellei niiden tukena käytetä kehittyneitä suodattimia. Monesti vuorovaikutus on loukkaavaa myös toiseen suuntaan, käyttäjältä kielimallille. Esimerkiksi kalifornialaisen Pandorabotsin toimitusjohtaja Lauren Kuntze on todennut, että noin kolmannes miesten kirjoittamista viesteistä yrityksen palkitulle chatbotille Mitsukulle on loukkaavia, seksuaalisesti häiritseviä tai vihjailevan romanttisia⁵⁶.

Seksuaalista häirintää voi estää suodattimilla, mutta niiden käytössä on monenlaisia haasteita. Osuva esimerkki vääristyneeseen aineistoon liitetystä sukupuolisesta stereotyyppiästä on sana ”lesbo”. Se yhdistyy valtavassa osassa verkkosisältöjä pornoteollisuuteen eikä sukupuoli-identiteettiin. Tästä syystä sen käyttö jälkimmäisessä merkityksessä jää usein suodattimien hampaisiin sosiaalisessa mediassa ja voi johtaa jopa sensuuriin ja julkaisukielloihin⁵⁷.

Generatiivista tekoälyä voidaan käyttää myös verkkovainossa ja häirinnän kohdistamisessa tiettyihin henkilöihin tai ryhmiin. Sen avulla voidaan luoda erittäin vakuuttavia tekaistuja viestejä tai profiileja, joita käytetään tietojen urkkimiseen tai luottamuksen rakentamiseen ennen häirintää. Niin kutsutuilla syvävääreännöksillä (deepfake) taas voidaan tehdä tekaistuja kuvia tai videoita ihmisistä. Tekoäly luo uusia tapoja syrjiä, loukata tai saattaa ihmiset ja ihmisryhmät huonoon valoon, esimerkiksi algoritmien kuplauttavan vaikutuksen tai generatiivisella tekoälyllä tuotetun misogynian kautta. Toisaalta tekoälyä voidaan käyttää myös sukupuolittuneen tekstin, stereotyypioiden ja jopa misogynian automaattiseen tunnistamiseen. Teknologiavälitteinen ja tekoälyä hyödyntävä naisiin kohdistuva vaino ja häirintä ovat kansainvälisesti laajaa huomiota saaneita ilmiöitä. Muun muassa Euroopan neuvosto alainen GREVIO-asiantuntijaryhmä määrittelee tekoälyä hyödyntävän naisiin kohdistuvan vainon ja häirinnän osaksi naisiin kohdistuvan väkivallan ja lähisuhdeväkivallan jatkumoa⁵⁸. Haasteiden tunnistamisen avulla voidaan mahdollisesti myös torjua niin teknologian kuin yhteiskunnan luomia eriarvoisuuksia ja syrjintää.

⁵⁵ Fulgu & Capraro, 2024

⁵⁸ GREVIO, 2021

⁵⁶ Balch, 2020

⁵⁷ Bartoletti & Xenidis, 2023, s. 27

2.5 Ennustavan tekoälyn haasteet

Myös julkishallinnossa käytetään tekoälyä eri palveluiden parempaan kohdentamiseen ja työtehtävien tehostamiseen. Tekoälyjärjestelmien mahdollisuus ennakoida palvelutarpeita tietyllä alueella tai eri väestöryhmissä on herättänyt erityisen paljon kiinnostusta. Esimerkiksi sosiaali- ja terveydenhuollon palveluiden tarvetta voidaan ennustaa koneoppimismenetelmillä melko hyvin, ja tällöin resursseja voidaan kohdistaa paremmin ja oikeudenmukaisemmin. Koneoppimismallien haasteena on, että jos jonkin tauti- tai potilasryhmän hoito on aliresursoitu, koneoppimismalli oppii tämän ryhmän kustannukset pieniksi, eikä välttämättä lisää resursseja tarvitseväksi kohteeksi.

Seuraava esimerkki kuvaa, kuinka hyvä tarkoitus ja resursseissa säästäminen eivät riitä syiksi automatisoida päätöksentekoa tekoälyllä. Esimerkki osoittaa myös, kuinka sukupuolten tasa-arvo risteää etnisyyden ja yhteiskunnallisen aseman kanssa, eikä sitä voi tarkastella erossa niistä. Virginia Eubanksin⁵⁹ kuvaaman hankkeen toteutuspaikkana oli yhdysvaltalainen Allegheny County Office of Children, Youth, and Families läntisessä Pennsylvaniassa. Virastossa käytettiin algoritmista mallia lasten hyväksikäytön todennäköisyyden ennustamiseen. Tavoitteena oli löytää kriteerejä, joiden avulla tämä voidaan tehdä ennalta ehkäisevästi ja siten suojella lapsia. Hyvä tarkoitus kuitenkin johti monella tavoin syrjivään prosessiin. Toisin kuin monissa muissa tilanteissa, tällä kertaa marginalisoiduista ryhmistä oli tarjolla dataa huomattavasti muita ryhmiä enemmän, etenkin asioista, joihin algoritmien ennustukset nojasivat. Naiset, rodullistetut ja vähävaraiset olivat yliedustettuina kategorioissa, joista algoritmeille sopivaa valvontadataa kerättiin.

Näiden ryhmien edustajista oli saatavilla runsaasti dataa julkiselta puolelta; lasten sosiaalipalveluista, mielenterveyspalveluista ja päihdepalveluista, kun taas varakkaamilta tätä tietoa ei löytynyt, koska he käyttivät etupäässä yksityisiä palveluita. Julkisten palveluiden käyttäjät olivat tekoälyn käyttämässä analyysissä yliedustettuja ja täten myös heidän lapsiinsa kiinnitettiin enemmän huomiota. Tässä tapauksessa tekoälymalli – tilastomatematisesti täysin oikein, mutta vääristyneen aineiston perusteella – ennusti, että he ovat suuremmassa riskissä joutua hyväksikäytetyiksi. Tekoälyn käyttö tehosti tuntuvasti, analyysiä mutta tuotti vääristyneitä tuloksia, jotka muodostivat uhan muutenkin syrjityille ja tukea tarvitseville perheille. Tekoälyä hyödyntävät mallinnukset voidaan saada toimimaan tehokkaasti ja oikein, jos aineistot ovat edustavia ja niiden mahdolliset vinoumat ja suhde kulloiseenkin kontekstiin selvillä.

Palveluiden paremman kohdentamisen lisäksi tekoälyjärjestelmiä on myös rakennettu paljastamaan ja ennakoimaan etuuspetoksia. Alankomaissa vuonna 2019 paljastunut massiivinen skandaali, jossa lapsilisiin liittyvää vilppiä kartoitettiin algoritmien avulla, johti katastrofaalisiin seurauksiin – yli 50 000 perheeltä perittiin perusteetta takaisin tuchia ja yli 1 500 lasta otettiin huostaan⁶⁰. Tästä huolimatta samantapaisia järjestelmiä on otettu käyttöön myös Pohjoismaissa. Tanskassa ja Ruotsissa tehtyjen selvitysten mukaan tekoälytyökaluilla on käytännössä luotu EU:n tekoälyasetuksen kieltämä pisteytysjärjestelmä, jolla on syrjiviä vaikutuksia esimerkiksi vammaisiin, pienituloisiin ja maahanmuuttajiin⁶¹.

⁵⁹ Eubanks, 2018

⁶⁰ van Gerven, 2024

⁶¹ van Gerven, 2024; Amnesty International, 2024

Kuten esimerkit edellä osoittavat, koneoppimismalleja on houkuttelevaa käyttää yksilötason ennusteiden ja vilppien paljastamiseen, mutta ne toimivat hyvin lähinnä väestö- ja ryhmätasolla⁶². Ennusteita ja niihin perustuvia interventioita ei voi suoraviivaisesti soveltaa yksilöihin. Jos tekoälyjärjestelmiä käytetään vilppien ja väärinkäytösten tunnistamiseen, eli ihmisten valvomiseen ja rankaisemiseen, voi sillä olla negatiivisia vaikutuksia teknologian hyväksyttävyyteen sekä ihmisten luottamukseen hyvinvointivaltion instituutioita kohtaan.

2.6 Tekoäly ylläpitää, uusintaa ja vahvistaa yhteiskunnan eriarvoisuuksia

Tämän luvun esimerkit osoittavat, että tekoälyyn liittyvässä sukupuoleen perustuvassa syrjinnässä ei ole kyse pelkästään ohjelmointivirheistä, jotka voidaan ratkaista yksinkertaisilla korjaustoimenpiteillä. Ongelmien ratkaisemiseksi on sen sijaan tärkeää tunnistaa yhteydet teknologian ja yhteiskunnan välillä. Tekoäly uusintaa, ylläpitää ja vahvistaa yhteiskunnassa olevia rakenteita ja vinoumia. Syrjivät ja sukupuolittuneet kielelliset ilmaisut, tiedonkeruu- ja luokittelutavat tai ajattelumallit siitä, mitä on maskuliinisuus tai feminiinisyys, heijastuvat tekoälyn käyttämään aineistoon sekä tekoälyjärjestelmien päätelmiin⁶³.

Tekoälyjärjestelmien arvioimisessa, kehittämisessä ja käyttämisessä tulee huomioida sosiaalinen, yhteiskunnallinen tai sektorikohtainen konteksti, jossa järjestelmää käytetään. Kontekstin huomioiminen tarkoittaa tarkkuuden ja herkkyyden lisäämistä virhetulkintojen välttämiseksi. Vaikka tämä vaatii enemmän resursseja, tulokset ovat silloin täsmällisempiä ja helpommin yhteiskunnallisesti hyväksyttäviä.

Usein keskitytään ainoastaan oireiden lievittämiseen, kuten syrjään jäävien ryhmien edustavuuden lisäämiseen ohjelmoinnin avulla tai loukkaavan sisällön sensurointiin suodattimilla, mutta juurisyihin ei puututa. Yksinkertainen syötteiden ja suoritteiden (inputs and outputs) säätäminen ei ole tehokas tai riittävä menetelmä, jos ongelmien perimmäiset syyt ovat vallan epätasapainossa ja yhteiskunnallisessa kontekstissa. Sukupuolten tasa-arvon näkökulmasta pinnalliset teknologiset korjaustoimenpiteet ovat riittämättömiä tai jopa haitallisia⁶⁴ ja voivat ylläpitää vaikutelmaa, että asia otetaan vakavasti ja siihen puututaan. Hienosäädön sijaan tarvitaan järjestelmällistä uudistamista, dataetiikkaa ja dataoikeudenmukaisuutta, joka pureutuu ongelmien perusteisiin saakka.

Kirjassaan *Data Feminism*⁶⁵ Catherine D'Ignazio ja Lauren F. Klein hahmottelevat jakoja teknisten korjausliikkeiden (vasemmalla) ja aidosti tasa-arvoa ajavien datajärjestelmien (oikealla) välillä. Vasemmalla ongelmien syyksi nähdään teknologia tai yksilöiden toiminta, kun taas oikealla keskitytään rakenteellisiin eroihin ja mahdollisuuksiin vaikuttaa niihin. Jakoa voidaan soveltaa myös tekoälyjärjestelmiin.

⁶² Salganik, 2020

⁶⁵ D'Ignazio & Klein, 2023, s. 60

⁶³ Bartoletti & Xenadis, 2023, s. 15

⁶⁴ Bartoletti & Xenadis, 2023, s. 32

Vinoumien korjaaminen

Aidosti tasa-arvoisten tekoälyjärjestelmien luominen

Läpinäkyvyys	Refleksiivisyys (mukautuminen)
Tasa-arvo (kaikille sama)	Oikeudenmukaisuus (kaikille tarpeiden mukaan)
Algoritmien ymmärtäminen	Yhteiskunnan, historian, kulttuurin ja kontekstin ymmärtäminen

Taulukko 1: Tekniset korjausliikkeet (vasemmalla) vai aidosti tasa-arvoa ajavat datajärjestelmät (oikealla)? Lähde: Data Feminism (D'Ignazio & Klein 2023).

Vinoumien ja vääristymien korjaaminen jälkikäteen eroaa huomattavasti tekoälyjärjestelmistä, joissa niiden vaikutuksiin on varauduttu jo edeltä käsin ja joissa niiden esiintyminen pyritään estämään rakenteellisesti, ei puuttamalla yksittäisiin tapauksiin. Vinoumien korjaaminen nähdään suoraviivaisena teknologisenä ratkaisuna, jolla rakennetaan tasa-arvoisempia tekoälyjärjestelmiä kärsivällisesti ja korjaus kerrallaan. Tällöin kuitenkin ollaan puuttumatta ongelmien juurisyihin, eikä korjausurakka pääty koskaan. Jos ennuste tekoälyn käytön lisääntymisestä yhä uusilla aloilla pitää paikkansa, myös vääristymiä tulee korjattavaksi valtavasti lisää. Tähän ei voi vaikuttaa muuten kuin puuttamalla rakenteisiin.

Läpinäkyvyys on termi, joka usein liitetään vastuulliseen ja ymmärrettävään tekoölyyn. Se tarkoittaa, että tekoälyjärjestelmien toiminta, päätöksentekoprosessit ja käytetyt algoritmit ovat ymmärrettäviä ja selkeästi kommunikoituja järjestelmän käyttäjille. Läpinäkyvyys voikin parantaa tekoälyn luotettavuutta, mutta ei valitettavasti tarjoa ratkaisua kaikkiin ongelmiin. Hannu Toivonen kuvaa läpinäkyvyyden säilyttämisen teknisiä haasteita:

”Kun tekoäly on ymmärrettävää, sen toimintaperiaatteet voidaan tarkistaa ja sen tekemät toimenpiteet perustella. Läpinäkyvyys voi mahdollistaa myös rikkinäisten kohtien korjauksen, jos toimenpiteet tai niiden perustelut ovat huonot. Läpinäkyvyyden vastakohta on musta laatikko, jonka sisään ei näe: tekoäly, joka näyttää testien perusteella toimivan hyvin, mutta kukaan ei pysty selittämään, miten... Miksei kaikista tekoölyistä sitten tehdä läpinäkyviä? Siksi, että joskus läpinäkymättömät ratkaisut, tyypillisesti neuroverkot, ovat läpinäkyviä huomattavasti parempia – kuten esimerkiksi monissa kuvankäsittelytehtävissä. Ja siksi, että ohjelman kehittäjälle voi olla helpompaa tuottaa koneoppimisella musta laatikko kuin miettiä itse, miten tehtävä kannattaisi ratkaista.”⁶⁶

⁶⁶ Toivonen, 2023, s. 73–74

Refleksiivisyys tarkoittaa tässä sitä, että jo suunnitteluvaiheessa pyritään luomaan järjestelmiä, joita voidaan muokata tarpeen mukaan. Tällöin läpinäkyvyys on tärkeää, mutta se ei ole riittävä lähtökohta tekoälyjärjestelmien kehittämiseen. Myös tasa-arvoa pitää tarkastella kriittisesti. Etenkin laadullisessa tutkimuksessa on yleistä tehdä ero tasa-arvon (equality) ja oikeudenmukaisuuden tai kohtuuden (equity) välille. Ensimmäisellä tarkoitetaan tilannetta, jossa kaikkia kohdellaan samalla tavalla ja yhteneväisesti. Jälkimmäisessä taas huomioidaan osallistujien erilaiset lähtökohdat ja reagoidaan niihin tasavertaisen kohtelun takaamiseksi. Tätä eroa voidaan havainnollistaa vammaisten henkilöiden näkökulmasta. Jos esimerkiksi digitaalinen työkalu, joka on välttämätön arjen sujuvuuden kannalta, on suunniteltu niin, ettei sitä voi käyttää näkövammaisena, pelkkä oikeus käyttää työkalua ei riitä takaamaan yhdenvertaisuutta. Näkövammaisille käyttäjille suunnittelemattomuus tarkoittaa tosiasiallisesti käytön estymistä, vaikka käyttöoikeus olisikin muodollisesti sama kaikille. Sukupuolten tasa-arvon suhteen nämä kysymykset ovat usein huomattavasti monimutkaisempia, ja tasa-arvoa täytyy pohtia historiallisessa kontekstissa ja ottaa huomioon valtaan liittyvät kysymykset⁶⁷. Edellä olevassa taulukko 1 erotteluista neljäs koskee teknologian paikkaa maailmassa. Algoritmit ovat kietoutuneet arkeemme yhä kasvavissa määrin, eikä teknologiaa voi tarkastella kattavasti omana erillisenä saarekkeenaan. Tekoälyn ymmärtäminen pelkkinä algoritmeina ja koodinpätkinä on riittämätön lähestymistapa, etenkin jos tarkoituksena on ottaa huomioon sen käytön sosiokulttuurisia ulottuvuuksia. Tekoäly toimii aina sidoksissa kontekstiinsa, joka heijastaa historiaa ja valtasuhteita.

⁶⁷ D'Ignazio & Klein, 2023, s. 62

3. Tekoälyn kehityksen vaikutus työelämän muutokseen

Tässä luvussa keskitytään tekoälyn ja työelämän yhtymäkohtiin. Teknologia-alan työpaikat ovat voimakkaasti sukupuolittuneet niin Suomessa kuin globaalisti⁶⁸ ja sama näkyvä tekoälyn kehittämisessä. Luvussa avataan myös rekrytointiin ja työn tehostamiseen liittyviä näkökulmia.

3.1 Tekoälyteollisuuden sukupuolijakauma

Tekoälysovelluksilla ja tekoälyn kehittämisellä on ollut monia vaikutuksia sukupuolten tasa-arvoon työelämässä. Tasa-arvokysymykset nousevat esiin heti, kun tarkastellaan, ketkä tekevät työtä tekoälysovellusten kehittämisen parissa. Kehitystyötä tehdään suurelta osin miesten johdolla, miesenemmistöisissä tutkimusorganisaatioissa ja yrityksissä. Maailmanlaajuisesti naisten osuus tekoälyyn liittyvissä työpaikoissa on kasvanut hitaasti. World Economic Forumin raportin⁶⁹ mukaan naisia on noin 30 % alan työntekijöistä. Teknologiateollisuuden⁷⁰ mukaan Suomessa kevään 2024 yhteishaussa tekniikan ja ICT-alan koulutukseen hakeneista 28 % oli naisia. Tutkimusten mukaan teknologia yhdistetään useammin miehiin kuin naisiin, mikä on aiheuttanut sen, että naiset väheksyvät ja aliarvoivat omia teknologiaan liittyviä taitojaan. Alan segregaaation lisäksi selvityksen kirjoittamisen tueksi järjestetyissä monitieteisissä ja moniammatillisissa pyöreän pöydän keskusteluissa nousi esiin, että myös tekoälyyn liittyvät teemat tai aiheet ovat usein sukupuolittuneita: naiset puhuvat etiikasta ja miehet teknologiasta.

Toisaalta voidaan kysyä, miksi teknologian kehittämisen sukupuolittuneisuus on ongelmallista, vai onko se sitä. Aihetta voidaan lähestyä moninaisen osaamisen ja näkökulmien kautta. Tekoälyjärjestelmien kehittäjät ja käyttöönottajat eivät pysty aina tunnistamaan vinoumia datassa tai omia ennako-oletuksiaan, joita heillä on esimerkiksi ihmisten käyttäytymisestä. Usein valtaväestöön kuuluvan on vaikea hahmottaa esimerkiksi syrjintää, jota vähemmistöön kuuluva kokee. Jos kehitystä tekevä ryhmä on hyvin homogeeninen – eli koostuu samankaltaisista ihmisistä – voivat muunlaisten ihmisten ja erityisryhmien näkökulmat, piirteet ja tarpeet jäädä helposti huomioimatta⁷¹. Valaiseva esimerkki löytyy tekoälyavusteisesta puheentunnistuksesta. Alunperin puheentunnistusteknologiaa kehittivät miesenemmistöiset ryhmät, jotka käyttivät ainoastaan miesten puhetta opetusdatana. Sen seurauksena puheentunnistusohjelmat tunnistivat huonosti naisten puhetta sen erilaisen äänenkorkeuden takia.

⁶⁸ World Economic Forum, 2023

⁷¹ D'Ignazio & Klein, 2023

⁶⁹ World Economic Forum, 2023

⁷⁰ Teknologiateollisuus, 2024

3.2 Vinoumat rekrytoinnissa

Monimuotoisten tutkimus- ja kehitysryhmien on raportoitu olevan innovatiivisempia⁷². Monissa raporteissa⁷³ on tunnistettu, että teknologia-alalle ja erityisesti teknologian kehittämiseen tarvitaan naisia ja ylipäänsä lisää moninaisista taustoista tulevia ihmisiä. Onkin ehdotettu, että rekrytointiprosesseissa voitaisiin käyttää tekoälyä ja erityisesti kielimalleja apuna tukemaan tasa-arvoisempaa ja yhdenvertaisempaa työelämää. Oikein suunniteltuna ja koulutettuna voi auttaa vähentämään inhimillistä puolueellisuutta rekrytointiprosessissa. Toisaalta on syytä huomioida, että jos koulutusdata sisältää vinoumia ja puolueellisuuksia, voivat tekoälyavusteiset rekrytointityökalut vahvistaa niitä.

Ensimmäiset kokeilut tekoälyn hyödyntämisestä rekrytoinnissa tuottivat sukupuolen kannalta syrjiviä tuloksia. Esimerkiksi suuryritys Amazonin rekrytointiprosessin on todettu suosineen miehiä teknologisiin tehtäviin liittyneissä rekrytoinneissa. Tekoälyjärjestelmän koulutuksessa käytetyt tiedot koostuivat pääosin miespuolisten työntekijöiden ansioluettelosta, sillä useimmat näissä tehtävissä työskentelevistä olivat miehiä. Tällöin prosessi toisti vallalla olevaa sukupuolijakaumaa, eikä kyennyt muuttamaan sitä. Virheistä ja vinoumista on sen jälkeen opittu ja uusissa järjestelmissä pyritään välttämään tämänkaltaiset virheet, mutta kaikkia vinoumia ei ole helppo ennakoida etukäteen. Vuonna 2024 voimaan astunut EU:n tekoälyasetus, josta tarkemmin lisää luvussa 4, määrittelee rekrytointiin liittyvät tekoälyjärjestelmät suuren riskin järjestelmiksi, jolloin niiden vaikutukset tulee arvioida etukäteen esimerkiksi sukupuolten tasa-arvon kannalta.

3.3 Tekoäly muuttaa työelämää

Tekoälyratkaisuilla pyritään tehostamaan työntekoa ja lisäämään tuottavuutta. Tasa-arvon näkökulmasta on tärkeää pohtia, mitä työssä voidaan ja halutaan tehostaa, ja mihin työntekijäryhmiin tehostamispyrkimykset kohdistuvat. Yhtäältä on esitetty, että tekoälysovelluksilla voidaan korvata asiantuntijatyötä, kun taas toisaalta on arvioitu, että sen avulla voidaan erityisesti automatisoida vähän asiantuntemusta vaativat rutiinityöt⁷⁴. Tällä hetkellä on vielä vaikea ennakoida, ovatko jotkut tietyt työpaikat tai tehtävät erityisesti vaarassa⁷⁵. Kyse ei kuitenkaan ole ennalta määrätystä kehityskulusta, vaan suunnittelun ja päätösten seurauksista eli kehitykseen voidaan vaikuttaa sekä poliittisten että organisaatioiden päätösten kautta. Yksittäisten päätösten tai esimerkiksi investointien vaikutuksia eri ihmisryhmiin voidaan ja tulee myös ennalta-arvioida.

Teknologinen muutos hävittää yleensä yllättävän harvoin kokonaisia ammatteja. Useimmiten muutos koskee työn kuvaa ja tapahtuu tehtävien tasolla ja siinä, kuinka paljon työntekijöitä tarvitaan erilaisten tehtävien suorittamiseen⁷⁶. Tekoälytyökalujen käyttö tuo myös uusia tehtäviä työhön, jotka liittyvät esimerkiksi datan kirjaamiseen ja hallintaan. Tällaiset datatyötehtävät ilmaantuvat tyypillisesti aikaisempien tehtävien rinnalle ilman, että tätä otetaan täysin huomioon töiden mitoituksessa. Esimerkiksi lääkärit ja hoitajat ovat kritisoineet sitä, että datatyö vie aikaa itse hoitotyöltä. Lisäksi tutkimukset⁷⁷ ovat

⁷² Hewlett et al., 2013

⁷³ UNESCO, 2022; OECD, 2024

⁷⁴ Ylikoski, 2024

⁷⁵ Kauhanen & Rouvinen, 2024

⁷⁶ Ylikoski, 2024

⁷⁷ Alastalo & Lehto, (hyväksytyt).

myös tuoneet esiin sen, että vaikka datatyö tulee osaksi monia tehtäviä, se ei näy tehtävien tai ammattien teknisessä arvostuksessa tai palkkauksessa.

Suomessa työmarkkinat ovat yhä voimakkaasti jakautuneet sukupuolen mukaan niin kunta-alalla, valtiolla kuin yksityiselläkin sektorilla. Vaikka tekoälyn tuomaa kehitystä on vaikeaa ennustaa, on esitetty, että teollisuudessa ja logistiikassa, jotka ovat miesenemmistöisiä aloja, tekoäly tulee automatisoimaan monia tehtäviä ja vähentämään työpaikkoja. Toisaalta on myös esitetty, että monilla naisenemmistöisillä aloilla, kuten toimistotyössä ja sosiaali- ja terveydenhuollossa, tehostamistarpeet koetaan suurina ja tekoälysovellusten käyttöönottoa edistetään näillä aloilla siksi voimakkaasti. Jos teknologia-alalla toimivat miehet kehittävät tekoälysovelluksia naisenemmistöisille aloille, on paikallaan pohtia mahdollisia vinoumia tai puutteita järjestelmissä – tunnistetaanko esimerkiksi piilevä datatyö ja hahmotetaanko työnkuva kokonaisvaltaisesti? Kohdistuuko automatisointi työntekijöiden näkökulmasta oikeisiin työn vaiheisiin?

Suomalaisessa tutkimuksessa tekoälypohjaisesta ennakointityönkalusta lastensuojelussa⁷⁸ kävi ilmi, että organisaatiossa kokeiltu tekoälysovellus istui huonosti käytännön sosiaalityöhön. Ennakoiva työkalu teki riskiarvioita perustuen asiakkaiden dataan – sekä yksilön että koko asiakaskunnan. Sosiaalityöntekijät taas perustavat oman arvionsa asiakastyöhön, jossa keskitytään kokonaisvaltaisesti tunnistamaan asiakkaiden nykyhetken ongelmia ja syitä niihin. Tekoälytyökalu oletti samanlaisen historian tarkoittavan samanlaisia ongelmia tulevaisuudessa: jos tietyt riskitekijät ovat yleensä tarkoittaneet, että tietty ongelma toteutuu, se toteutuu myös tällä asiakkaalla. Sosiaalityöntekijät taas painottivat arvioissaan yksittäisen asiakkaan mahdollisuuksia vaikuttaa itseään koskevaan tilanteeseen.

Tekoälysovellusten kehityksessä ja konkreettisten työkalujen suunnittelussa keskeistä on ottaa mahdollisuuksien mukaan niin loppukäyttäjät kuin järjestelmien toimenpiteiden kohteet mukaan kehittämisprosessiin. Tätä painotetaan myös tekoälyyn liittyvissä eettisissä ohjeistuksissa sekä EU:n tekoälyasetuksessa. Samoin on syytä arvioida automatisoinnin ja tehostamisprosessien vaikutus ympäröivään yhteiskuntaan. Jos sairaanhoitajien tai teollisuuden töitä korvataan digitaalisilla ratkaisuilla ja tekoälyllä, on tarpeen pohtia, mitä se tarkoittaa eri työmarkkinoille ja erilaisten alueiden elinvoimaisuudelle. Minkälaisia lisä- tai uudelleen koulutuksen tarpeita ilmenee ja miten niihin voidaan varautua?

Koulutus ja osaaminen

Tasa-arvon ja työelämän kannalta tulee myös kiinnittää huomiota ihmisten kykyihin ja mahdollisuuksiin käyttää tekoälysovelluksia ja ymmärtää niiden toimintaperiaatteita, vinoumia ja rajoituksia. EU:n tekoälyasetuksen johdantolauseissa esitetään, että kaikilla tekoälyn kanssa toimivilla tulisi olla riittävä tekoälylukutaito⁷⁹:

⁷⁸ Lehtiniemi, 2023

⁷⁹ Tekoälyasetus, s. 6

”Jotta saadaan mahdollisimman paljon hyötyä tekoälyjärjestelmistä ja samalla suojellaan perusoikeuksia, terveyttä ja turvallisuutta ja mahdollistetaan demokraattinen valvonta, tarjoajilla, käyttöönottajilla ja henkilöillä, joihin vaikutukset kohdistuvat, olisi oltava riittävä tekoälylukutaito, jotta he ymmärtävät tarvittavat käsitteet ja voivat tehdä tekoälyjärjestelmiä koskevia tietoon perustuvia päätöksiä.”

Asetuksen mukaan organisaatioiden tulee huolehtia 2.2.2025 alkaen, että henkilöstö saa koulutusta tekoälyn hyödyistä ja riskeistä. Monet organisaatiot ovat ryhtyneet kehittämään ohjeistuksia työntekijöilleen. Esimerkiksi Kuntaliiton tutkimuksen mukaan⁸⁰ 75 % prosenttia kunnista oli jo ottanut käyttöön yleiset tekoälyyn liittyvät ohjeet tai oli kehittämässä niitä. Myös tekoälyn eettiseen käyttöön liittyviä ohjeita oli kyselyn mukaan olemassa jo monissa kunnissa, mutta sopimuksiin ja hankintoihin liittyviä ohjeita ei juurikaan ollut. Kyselyssä myös ilmeni, että monilla kunnilla oli vaikeuksia arvioida, minkälaista koulutusta tarvitaan.

Tekoälylukutaidon lisäksi puhutaan myös datalukutaidosta ja digitaalisesta lukutaidosta. Kaikkien näiden lukutaitojen kohdalla on kuitenkin tärkeä havaita, että ymmärtääkseen tekoälyä tai datafikoituvaa työelämää, teknisten taitojen lisääminen ei riitä⁸¹. Esimerkiksi rakenteellisten vinoumien ja mahdollisten syrjivien mekanismien tunnistaminen organisaation käyttämässä datassa ja järjestelmissä ei vahvistu vain lisäämällä koulutusta koodaukseen tai data-analytiikkaan. Sen lisäksi tarvitaan osaamista lainsäädännöllisistä ja eettisistä periaatteista sekä itse käsiteltävästä asiasta. On myös tärkeää hahmottaa ja tuoda esiin datatyöhön liittyvät tehtävät ja työmäärät, jotka jäävät usein piiloon.

Vaikka tekoälyn kaikkia vaikutuksia työelämään on vaikea ennustaa, kehitystä ohjaavilla poliittisilla ja työpaikkakohtaisilla päätöksillä on suuri merkitys siinä, mille aloille panostetaan tai mitä työtehtäviä halutaan automatisoida. Koska suomalaiset työmarkkinat ovat voimakkaasti jakautuneet naisten ja miesten aloihin ja tehtäviin, tekoälyn vaikutukset työelämään ovat myös todennäköisesti sukupuolittuneita.

⁸⁰ Kuntaliitto, 2024

⁸¹ Ylikoski & Snell, 2021

4. Kansainvälinen ja kansallinen tekoälyyn liittyvä sääntely

Tässä luvussa esitellään lyhyesti sukupuolten tasa-arvoon liittyvää sääntelykehikkoa. Lisäksi luvussa nostetaan esiin muutamia keskeisiä perusoikeuksiin, syrjintään ja tasa-arvoon liittyviä näkökulmia, joita on kirjattu 2024 hyväksytyyn Euroopan unionin tekoälyasetukseen sekä erilaisiin tekoälyn etiikkaa koskeviin periaatteisiin. Luvun lopuksi esitellään tekoälysääntelyyn liittyvää keskustelua ja kriittisiä näkökulmia.

4.1 Sukupuolten tasa-arvoon liittyvä sääntely ja tekoäly

Teknologian ja tekoälyn ihmisvaikutuksiin liittyvän sääntelyn taustalla on joukko ihmisten perusoikeuksiin liittyviä sopimuksia ja säädöksiä, jotka tulee ottaa huomioon tekoälyn käytössä. Suomen kansallisessa lainsäädännössä, Euroopan unionin lainsäädännössä ja useissa Suomea sitovissa kansainvälisissä sopimuksissa säännellään sukupuolten tasa-arvon edistämisen velvoitteista, yhdenvertaisesta kohtelusta ja syrjinnän ehkäisystä.

Suomen perustuslain mukaan ihmiset ovat tasa-arvoisia lain edessä. Perustuslain mukaan ketään ei saa ilman hyväksyttävää syytä asettaa eri asemaan sukupuolen, iän, alkuperän, kielen, uskonnon, vakaumuksen, mielipiteen, terveydentilan, vammaisuuden tai muun henkilöön liittyvän syyn perusteella. Perustuslaki velvoittaa myös edistämään sukupuolten tasa-arvoa yhteiskunnallisessa toiminnassa ja työelämässä erityisesti palkkauksen ja muiden palvelussuhteen ehtojen osalta. Tämän lisäksi sukupuolten tasa-arvon edistämistä säätelee Suomessa tasa-arvolaki⁸². Tasa-arvolain tarkoituksena on estää sukupuoleen perustuva syrjintää ja edistää sukupuolten tasa-arvoa.

Myös useat EU:n direktiivit ohjaavat kansallista sukupuolten tasa-arvoon linkittyvää lainsäädäntöä. Niin kutsutut tasa-arvodirektiivit koskevat muun muassa työelämän tasa-arvoa, samapalkkaisuutta ja vanhempainvapaita. Vuonna 2024 hyväksytty direktiivi naisiin kohdistuvan väkivallan ja perheväkivallan torjumisesta kattaa muun muassa verkkoväkivallan eri muodot. Direktiivin johdanto-osiossa mainitaan myös tekoäly. EU:n direktiivien lisäksi keskeinen ohjaava vaikutus on niin EU:n tuomioistuimen oikeuskäytännöllä kuin EU:n tasa-arvostrategialla.

⁸² Laki naisten ja miesten välisestä tasa-arvosta, 608/1986

Tasa-arvon edistämiseen ja syrjinnän ehkäisyyn velvoittavat myös kansainväliset sopimukset. Yksi keskeisiä kansainvälisiä tasa-arvosopimuksia on YK:n Kaikkinaisen naisten syrjinnän poistamista koskeva yleissopimus (CEDAW), jonka tavoitteena on edistää ihmisoikeuksien tasa-arvoista toteutumista naisten osalta.

Kuten selvityksessä on todettu, tekoäly tuo uudenlaisia haasteita siihen, miten tunnistamme ja ymmärrämme sukupuolten tasa-arvon ja kuinka tekoälyjärjestelmät saattavat ylläpitää ja vahvistaa eriarvoisuutta. Muuttuva toimintaympäristö voi myös vaatia tasa-arvoon liittyvien säädösten ja sopimusten uudelleentarkastelua sekä uudenlaista toimintakulttuuria, jonka avulla data ja tekoäly voidaan valjastaa edistämään tasa-arvoa. YK:n kestävä kehityksen tavoitteet tarjoavat esimerkin siitä, miten datankeruuta ja tekoälyä voidaan hyödyntää tasa-arvon edistämiseksi maailmanlaajuisesti. Yksi Agenda2030-toimintaohjelman tavoitteista on saavuttaa sukupuolten tasa-arvo (tavoite 5). Eri puolilla maailmaa on ryhdytty panostamaan tähän tavoitteeseen liittyvään datankeruuseen ja tutkimukseen, jotta ongelmia olisi helpompi tunnistaa ja niille löytää ratkaisuja. Tekoälyä voidaan hyödyntää tämän datan analyysissä, mikä voi osaltaan auttaa edistämään tasa-arvoa.

Suomessa esimerkiksi yhdenvertaisuusvaltuutettu on esittänyt syrjinnän ehkäisyyn ja ihmisten perusoikeuksien edistämiseen liittyviä huomioita ja suosituksia tekoälyn hyödyntämisessä ja käyttöönottossa. Yhdenvertaisuusvaltuutettu muun muassa suosittaa, että valtionhallinnossa jalkautetaan arviointikehikko syrjimättömän tekoälyn varmistamiseksi. Arviointikehikko julkaistiin algoritmiseen syrjintään ja yhdenvertaisuuteen keskittyneen raportin osana vuonna 2022 yhdenvertaisuuden ja sukupuolten tasa-arvon edistämiseksi.⁸³

4.2 EU:n tekoälysäädös ja muu kansainvälinen sääntely

Euroopan unioni toimii maailmanlaajuisena edelläkävijänä ryhtyessään valmistelemaan tekoälyasetusta (AI Act), joka tuli voimaan vuonna 2024. Asetus otetaan jäsenvaltioissa käyttöön asteittain. Sen tarkoituksena on varmistaa, että EU:n alueella markkinoille tuotavat ja käyttöön otettavat tekoälyjärjestelmät eivät vaaranna ihmisten terveyttä, turvallisuutta tai perusoikeuksia. Asetus on riskiperusteinen, eli siinä säännellään tekoälyjärjestelmiä niiden aiheuttamien riskien perusteella. Tekoälyjärjestelmät on luokiteltu neljään eri tasoon, ja erittäin haitallisiksi luokitellut tekoälyn käyttötavat on kielletty kokonaan. Näitä ovat sovellukset, jotka voivat uhata kansalaisten oikeuksia. Kiellettyjä ovat esimerkiksi sosiaaliseen pisteytykseen liittyvät tekoälyjärjestelmät tai työpaikalla käytettävät tunteentunnistussovellukset.

⁸³ Yhdenvertaisuusvaltuutettu, 2022a; 2022b; Ojanen et al, 2022

Sosiaalinen pisteytys. Sosiaalisessa pisteytyksessä yksilöille tai organisaatioille annetaan pisteitä tai arvosanoja heidän käyttäytymisensä ja toimintansa perusteella. Pisteytys voi vaikuttaa ihmisten mahdollisuuksiin ja oikeuksiin yhteiskunnassa. Yksi tunnetuimmista esimerkeistä on Kiinassa 2020 käyttöön otettu sosiaalinen luottoluokitusjärjestelmä, jossa kansalaisten toimista kerätään tietoa ja ne arvotetaan⁸⁴. Esimerkiksi täsmällisyys laskujen maksussa tuottaa pluspisteitä, kun taas maan poliittisen järjestelmän arvostelusta voi saada miinus pisteitä. Jos kansalainen kerää liikaa miinus pisteitä, häneltä voidaan evätä lupa matkustaa ulkomaille tai hänen lapseltaan opiskelupaikka. Pisteytyksen pohjana olevassa ohjekirjelmässä⁸⁵ ei kuitenkaan avata, miten arvottaminen tehdään, joten varsinaisia päätöksiä täsmennetään aluetasolla. Järjestelmä on herättänyt paljon keskustelua ja huolta yksityisyyden suojasta ja ihmisoikeuksista.

Suuren riskin tekoälysovelluksille, kuten sellaisille tekoälyjärjestelmille, jotka koskevat lainvalvontaa, rekrytointia, laina-arviointia ja biometristä todentamista, on asetettu tiukennettuja vaatimuksia liittyen esimerkiksi avoimuuteen ja ihmisen suorittamaan valvontaan. Tekoälyasetus on lähtökohtaisesti osa markkinavalvontaa ja tuoteturvallisuuslainsäädäntöä, mutta tämän lisäksi säädös velvoittaa tekoälyjärjestelmien käyttöönottajia tekemään perusoikeuksiin liittyvän vaikutustenarvioinnin. Sukupuolten tasa-arvon kannalta tekoälyasetuksessa on keskeistä juuri perusoikeuksiin liittyvien riskien ennaltaehkäisy.

Sukupuolten tasa-arvo nousee tekoälyasetuksessa esiin myös monessa viittauksessa syrjimättömyyteen ja monimuotoisuuteen. Asetuksen johdantolauseissa todetaan esimerkiksi: ”Monimuotoisuudella, syrjimättömyydellä ja oikeudenmukaisuudella tarkoitetaan, että tekoälyjärjestelmiä kehitetään ja käytetään tavalla, jossa on mukana monimuotoisia toimijoita ja jolla edistetään yhtäläisiä mahdollisuuksia, sukupuolten tasa-arvoa ja kulttuurien monimuotoisuutta ja vältetään unionin tai kansallisessa lainsäädännössä kiellettyjä syrjiviä vaikutuksia ja epäoikeudenmukaisia vääristymiä⁸⁶.”

Tekoälyasetusta voidaan pitää hyvänä ensiaskeleena: sen tarkoitus on minimoida perusoikeuksiin liittyvät riskit. Asetusta on kuitenkin kritisoitu siitä, että perusoikeudet eivät näy siinä kovin konkreettisesti⁸⁷.

Tekoälyn tasa-arvoisen kehityksen edistämiseksi ja sääntelyssä sekä sen toimenpanossa on syytä pohtia:

- Miten perusoikeuksien, kuten tasa-arvon toteutumista sovelletaan ja tulkitaan konkreettisesti tekoälyasetuksen ja muiden säädösten puitteissa?
- Miten erilaiset moniperustaiset ja risteävät syrjinnän muodot voidaan ottaa huomioon nykyisen sääntelyn piirissä?
- Miten riskien ehkäisemisen lisäksi tasa-arvon ja yhdenvertaisuuden edistäminen saadaan osaksi tekoälysovelluksia?

⁸⁴ Helsingin Sanomat, 2020

⁸⁶ Tekoälyasetus, 2024, s. 8

⁸⁵ China Law Translate, 2015

⁸⁷ Lindroos–Hovinheimo, 2024

Tekoälyasetus määrittää korkean riskin sovellukset, joiden käyttöönnotossa on erilaisia toimenpiteitä ja vaatimuksia, joissa tulee arvioida tasa-arvon ja syrjimättömyyden toteutumista. Nämä arvioinnit on hyvä ulottaa kaikkiin tekoälyjärjestelmiin. Asetuksessa todetaan, että jäsenvaltioiden on helpotettava käytännösääntöjen laatimista, jotka koskevat tällaisten erityisvaatimusten vapaaehtoista soveltamista kaikkiin tekoälyjärjestelmiin – ei vain suuririskisiin. Tavoitteena olisi tällöin tukea kaikkien tekoälyjärjestelmien suunnittelua niin, että se olisi osallistavaa ja monipuolista. Tämä tapahtuisi muun muassa perustamalla osallistavia ja monimuotoisia kehitysryhmiä eri organisaatioihin ja edistämällä sidosryhmien osallistumista kehitysprosessiin. Lisäksi asetus kannustaa siihen, että kaiken tason tekoälyjärjestelmien kohdalla arvioitaisiin ja ehkäistäisiin niiden kielteisiä vaikutuksia haavoittuvassa asemassa oleviin henkilöihin sukupuolten tasa-arvon osalta.

Keskeisiä näkökulmia tekoälyasetuksen kansalliseen toimeenpanoon ovatkin siis yhtäältä perusoikeusriskien minimointi kaikissa tekoälyjärjestelmissä ja toisaalta se, miten voidaan edistää tasa-arvoa edistäviä järjestelmiä ja sovelluksia.

Tekoälyasetus on globaalisti merkittävä askel kohti tekoälyn sääntelyä. Euroopan neuvostossa hyväksyttiin vuonna 2024 tekoälyä koskeva puiteyleissopimus, joka on periaatteiltaan yhdenmukainen EU:n tekoälyasetuksen kanssa. Puitesopimuksen lähtökohtana on ihmisoikeuksia kunnioittava tekoälyn käyttö. Sopimus on oikeudellisesti sitova väline, jonka tarkoituksena on varmistaa, että tekoäly noudattaa yhteisiä ihmisoikeuksia, demokratiaa ja oikeusvaltiota koskevia periaatteita.

EU:n tekoälyasetus ei kuitenkaan sääntele kaikkea tekoälyä, kuten tekoälyjärjestelmien käyttöä kansallisen turvallisuuden ja puolustuksen tarpeisiin tai tieteellistä tutkimusta. Se ei myöskään syrjäytä muuta lainsäädäntöä, joka jo ohjaa teknologian ja datan käyttöä eri näkökulmista ja eri sektoreilla. Tällainen keskeinen elementti säädösympäristössä on EU:n yleinen tietosuojasetus eli GDPR. Se luo pohjan kaikelle muulle eurooppalaiselle digisääntelylle ja määrittää mahdolliset henkilötietojen käsittelyperusteet sekä kansalaisten oikeudet oman datansa suhteen myös tekoälyyn liittyen. Tekoälyn käytön ja sukupuolten tasa-arvon näkökulmasta merkityksellisiä ovat myös erilaiset ihmisten perusoikeuksiin ja ihmisoikeuksiin liittyvät sopimukset ja säädökset, kuten EU:n perusoikeuskirja ja Euroopan ihmisoikeussopimus, Euroopan neuvoston yleissopimus naisiin kohdistuvan väkivallan ja perheväkivallan ehkäisemisestä ja torjumisesta (Istanbulin sopimus) sekä YK:n Kaikkinaisen naisten syrjinnän poistamista koskeva yleissopimus (CEDAW).

Myös tekoälyn eettisiä ohjeistuksia on luotu kansainvälisesti ja EU:ssa. Euroopan unionin riippumaton tekoälyä käsittelevä korkean tason asiantuntijaryhmä julkaisi vuonna 2019⁸⁸ luotettavaa tekoälyä koskevat eettiset ohjeet. Asiantuntijaryhmä määritteli neljä eettistä periaatetta, joita tulisi noudattaa tekoälyn kehittämisessä ja käytössä:

1. Ihmisen itsemääräämisoikeuden kunnioittaminen,
2. vahinkojen välttäminen,
3. oikeudenmukaisuus sekä
4. selitettävyyys.

⁸⁸ Tekoälyä käsittelevä korkean tason asiantuntijaryhmä, 2019

Selitettävyyden tarkoittaa, että ihmiset voivat ymmärtää ja jäljittää tekoälyjärjestelmän tekemät päätökset. Päätöksenteossa käytettävien tekoälymallien läpinäkyvyys ja selitettävyyden ovat ensisijaisia tärkeitä oikeusvaltiossa. Lisäksi asiantuntijaryhmä esitti seitsemän vaatimusta, joiden kautta eettiset periaatteet voidaan muuttaa käytäntöön. Niihin kuuluivat esimerkiksi monimuotoisuuden, syrjimättömyyden ja oikeudenmukaisuuden edistäminen, sisältäen esimerkiksi vääristymien välttämisen ja sidosryhmien osallistumisen.

UNESCO julkaisi vuonna 2021⁸⁹ oman eettisen ohjeistuksensa, joka painottaa ihmis-
oikeuksia ja ihmisarvoa. Eettisten periaatteiden lisäksi ohjeistuksessa identifioidaan yksitoista politiikka-
aluetta, joilla jäsenvaltioiden tulisi tehdä poliittisia toimenpiteitä vastuullisen tekoälyn kehittämiseksi. Yksi nimetyistä alueista oli sukupuoli. Riskien tunnistamisen ohella suosituksissa painotetaan, että jäsenmaiden tulisi edistää sellaisia tekoälyjärjestelmiä, jotka mahdollistavat tasa-arvon saavuttamisen. Vaikka suositukset eivät sido jäsenvaltioita tai tekoälyä kehittäviä yrityksiä, ne ohjaavat osaltaan tekoälystä käydyn keskustelun suuntaa.

4.3 Sääntelyn haasteet ja mahdollisuudet

Tekoälyn ja sukupuolten tasa-arvon edistämiseksi sääntelykysymykset koskevat laajaa kenttää toimijoita, jotka tulevat eri aloilta, käyttävät eri terminologioita ja lähestyvät aihetta eri lähtökohdista. Tekoälyn sääntelyyn liittyvät kysymykset herättävät myös paljon keskustelua asiantuntijapiirien ulkopuolella ja nousevat säännöllisesti esille tiedotusvälineissä. Julkisessa keskustelussa esiintyy yhtäältä väittämiä, että Euroopassa säädellään tekoälyä jo liikaa ja toisaalta, että sääntely ei pysy teknologian kehittämisen perässä⁹⁰. Lisäksi esimerkiksi tekoälyasetus sisältää erilaisia teknisiä termejä, mutta myös yleiskielessä käytettyjä, kuten haitta ja vahinko, joiden tulkinta voi olla hankalaa⁹¹.

Sääntelyä koskevat käsitteet eivät käänny helposti eri alojen välillä. Esimerkiksi vinoumia, läpinäkyvyyttä ja reiluutta on vaikea soveltaa yksityiskohtaisesti lain ja etiikan aloilla. Tämän vuoksi erityisalojen sisäiset keskustelut ovat erittäin tärkeitä ja kehittävät yksityiskohtaista tietoa asiantuntijoiden käyttöön. Ne kuitenkin vastaavat vain harvoin tekoälyn ja sukupuolten tasa-arvon kaltaisiin kysymyksiin, jotka vaativat tuekseen monitieteistä tutkimusta ja tarkastelua, joka huomioi moninaisia risteäviä näkökulmia.

Tekoälyn etiikkaan keskittyneet datatieteilijät Bartoletti ja Xenidis esittävät⁹², että kattavan syrjinnän vastaisen strategian täytyy ottaa huomioon ihmisoikeuksien ja muun lainsäädännön lisäksi algoritmisten päätöksentekojärjestelmien kaikki vaiheet kysymyksenasettelusta toimintakontekstiin, ja prosessin varsinaisesta suorittamisesta sen vaikutuksiin. Muuten vaarana on, että rajatut yksittäiset korjausliikkeet eivät osu kohteeseensa ja kuva kokonaisuudesta katoaa. Tekoälyasetus pyrkii osittain vastaamaan tähän haasteeseen painottamalla vaikutustenarviointia ja tekoälyjärjestelmien riskien ennakoimista myös matalariskisempien järjestelmien kohdalla. Uhkana kuitenkin on, että

⁸⁹ Unesco, 2022

⁹² Bartoletti & Xenidis, 2023

⁹⁰ Viljanen, 2023

⁹¹ Lindroos-Hovinheimo, 2024

riskejä arvioidaan valmiiden tarkistuslistojen kautta ilman refleksiivistä yhteiskunnan rakenteisiin liittyvää pohdintaa. (Katso taulukko 1 luvussa 2.6)

Tekoälyyn liittyvää sääntelyä on luotu myös paikallistasolla. Esimerkiksi New Yorkin kaupunginvaltuusto (City Council) otti käyttöön vuonna 2018, ensimmäisenä Yhdysvalloissa, laillisen sääntelyn algoritmeista johtuvaa syrjintää vastaan. Tämän myötä kaupunkiin muodostettiin myös työryhmä, joka keskittyy oikeudenmukaisuuden ja reilun toteutumiseen kaupungin käyttämissä algoritmeissa⁹³. Tämä on pelkkää teknologisten ominaisuuksien sääntelyä selvästi kokonaisvaltaisempi tapa toimia ja ottaa huomioon sosiokulttuurisen kontekstin, jossa tekoälyä käytetään.

Tekoälyyn perustuvilla menetelmillä voi tehdä tarkkanäköisiä analyyskejä suurista aineistoista erittäin tehokkaasti ja hahmottaa monimutkaisia kokonaisuuksia uusilla tavoilla. Niiden vastuullista käyttöä tulee rohkaista – mahdollisuuksia siihen löytyy jatkuvasti lisää. Sääntely voidaan kohdistaa myös vastuullisuuden edistämiseen ja löytää näin tekoälylle uusia, tasa-arvoa edistäviä käyttötarkoituksia.

Vastuukysymykset ovat tekoälyn kannalta keskeisiä, ja myös kiisteltyjä. Kuka on oikeudellisessa vastuussa monimutkaisen tekoälyjärjestelmän tuottamista vahingoista: kehittäjä, käyttöönottaja vai itse järjestelmä? Eräs tapa ajatella vastuuta on, että ihmisillä tulisi olla lopullinen kontrolli järjestelmän tuotoksista ja ihmisten tulisi olla aina mukana ainakin tarkistamassa tekoälyn päätelmiä (human-in-the loop). Tämä ei ole välttämättä helppoa järjestelmissä, jotka eivät ole läpinäkyviä. Lisäksi automatisaatioon yhdistetty ihmiskontrolli ei ole aina hyvä asia. Ihmisistä voi tulla huolimattomia ja välinpitämättömiä, jos heidän kontrollitehtävistään tulee liian rutiininomaisia⁹⁴.

Tekoäly-yritysten näkökulmasta teknologiaan voi myös liittyä liikesalaisuuksia, joita yritykset eivät mielellään luovuta ulkopuolisille, eivät välttämättä edes viranomaisille. Tilanteissa, joissa on syytä epäillä tekoälyn käytöstä johtuvaa syrjintää, on kuitenkin varmistettava, että viranomaisten tiedonsaantioikeus toteutuu. Tiedon saamisen lisäksi viranomaisilla tulisi olla riittävät resurssit ja osaaminen tulkita saamaansa aineistoa syrjintäepäilyn ratkaisemiseksi. Haasteena onkin juuri tämän resursoinnin ja osaamisen varmistaminen.

Tasa-arvon ja syrjimättömyyden kannalta tulee myös huomioida kansalaisten rooli tekoälyn käyttäjinä ja kohteina. Yksittäisten kansalaisten voi olla erittäin vaikea tunnistaa joutuneensa syrjinnän kohteeksi, koska tieto tekoälyn toimintaperiaatteista sekä lainsäädännöstä on vähäistä ja monille kokonaan tuntematonta. Lisäksi ihmisillä ei usein ole tarpeeksi tietoa omista oikeuksistaan henkilökohtaisten tietojensa käsittelyn suhteen. On myös vaikea tietää, miten todisteita algoritmien ja tekoälyn tuottamasta syrjinnästä tulisi tai voi kerätä⁹⁵. Samaan aikaan on kuitenkin olemassa erilaisia kansalaisjärjestöjä ja aktivistiryhmiä, jotka yrittävät nostaa tietoisuutta datasta, tekoälystä ja kansalaisten oikeuksista. Kansalaisia ei tulekaan ajatella vain tekoälyn kohteina tai algoritmisen syrjinnän uhreina vaan toimijoina teknologisoituneessa yhteiskunnassa.

⁹³ D'Ignazio & Klein, 2023, s. 58

⁹⁴ Viljanen, 2023

⁹⁵ Orwat, 2024

Data-aktivismi ja tekoälyn tulevaisuudet

Tekoälyn sääntely, kehittäminen ja käyttöönotto liittyvät myös valtaan ja valtarakenteisiin. Näitä kysymyksiä lähestytään dataoikeudenmukaisuutta ja data-aktivismia käsittelevissä puheenvuoroissa, joiden tarkoituksena on vaikuttaa suoraan syrjiviin valtarakenteisiin ja tarkastella kriittisesti olemassa olevia eettisiä ohjeita ja läpinäkyvyyden vaatimuksia⁹⁶. Suuri osa tekoälyn virallisen tason sääntelystä perustuu data-aktivistien alunperin esittämiin huomioihin ja epäkohtiin.

Data-aktivistit ovat viime vuosina nostaneet esiin dataoikeudenmukaisuuden kysymyksiä⁹⁷, jotka liittyvät esimerkiksi eri väestöryhmien oikeuksiin ja mahdollisuuksiin vaikuttaa heistä kerättävään tietoon ja sen käyttöön. Ruohonjuuritason aktivismi näkyy muun muassa erilaisten julistusten ja manifestien muodossa. Kantaaottavat tekstit tarjoavat erinomaisen näkymän tekoälyn käytön ongelmiin, sillä ne kokoavat yhteen keskeisiä aiheita datan ja etiikan solmukohdista ja vaativat oikeudenmukaisempia tulevaisuuksia.

Useissa data-aktivistien kannanotoissa, kuten Digital Justice Manifesto⁹⁸ ja Feminist Data Manifest-No⁹⁹ -manifesteissa, korostetaan tarvetta työskennellä aktiivisesti yhteiskunnan eriarvoisuuksien kitkemiseksi. Näistä jälkimmäisessä painotetaan, että dataan liittyvät riskit eivät ole samanlaisia kaikille ja kaikkialla eikä dataan perustuvaa valvontaa tulisi normalisoida, koska se kohdistuu eri tavoin eri sukupuoliin ja erilaisiin väestöryhmiin. Usein aktivistien kannanotoissa vaatimukset koskevat tiedollista oikeudenmukaisuutta. Tämä viittaa ajatukseen siitä, ettei riitä, että annetaan ääni marginaalissa oleville ryhmille, vaan että aktiivisesti työskennellään ihmisten oikeuksien puolesta ja eriarvoisuutta vastaan¹⁰⁰. Data-aktivistien ehkä keskeisin argumentti keskusteluissa digitaalisten tulevaisuuksien sääntelystä koskee kriittisen ajattelutavan vaalimista. Data-aktivistit korostavat, että digitaaliset ja yhteiskunnalliset tulevaisuudet voidaan myös kuvitella toisin ja niihin voidaan vaikuttaa. Tästä syystä heidän kannanottonsa auttavat kyseenalaistamaan teknologisten kehityskulkujen vääjämättömyyden.

Data-aktivistit korostavat esimerkiksi ihmisten tasapuolista oikeutta omaan dataansa: tällöin päämääränä on mahdollistaa digitaalisten tulevaisuuksien omistajuus¹⁰¹. Aktivistit vaativat, että kansainvälisen ihmisoikeusjärjestelmän pitää tunnistaa yhteys ihmisten ja heidän datansa välillä ja ymmärrys datasta taloudellisena resurssina, joka usein viedään yksilöiltä heidän käyttämiensä palveluiden omistajille¹⁰². Tämän kaltainen laajempi keskustelu oikeuksista dataan ei keskity vain henkilötietojen tai anonymiteetin turvaamiseen, vaan suurempaan kokonaisuuteen taloudellisia ja sosiaalisia oikeuksia¹⁰³. Ihmisten pitää pystyä ymmärtämään heistä kerätyn datan laajuus ja tekoälyn rakenteet¹⁰⁴.

On selvää, että henkilöä itseään koskevan datan kontrollia varten tarvitaan työkaluja, mutta on myös yhtä oleellista tähdentää aitojen vaihtoehtojen tarpeellisuutta datavetoisten ja tekoälyä hyödyntävien tuotteiden ja alustojen kanssa. Data-aktivistit esittävät lisäksi, että yksilöillä tulisi myös olla mahdollisuus ja oikeus kieltäytyä tiettyjen palvelujen käytöstä. Nykytilanteessa ei ole mahdollista jättäytyä kokonaan tekoälyä käyttävien

⁹⁶ Cifor et al., 2019

⁹⁷ ks. Puukko, 2021

⁹⁸ Just Net Coalition, 2019

⁹⁹ Cifor et al., 2019

¹⁰⁰ Milan & Treré, 2019

¹⁰¹ Just Net Coalition, 2019, s. 6

¹⁰² Just Net Coalition, 2019

¹⁰³ Puukko, 2021, s. 147–148

¹⁰⁴ Cifor et al., 2019

syrijvien rakenteiden ulkopuolelle ilman merkittäviä kielteisiä seurauksia. Esimerkiksi jättäytyminen pois sosiaalisen median alustoilta on erittäin vaikeaa, koska suuri osa vuorovaikutuksesta ja tiedotuksesta on keskittynyt niille. Dataoikeudenmukaisuuden periaatteet ohjeistavat kohti demokraattista digitaalista infrastruktuuria, joka toimii globaalilla tasolla. Tavoitteena on määrittää yksilöiden digitaaliset oikeudet sekä yritysten ja valtioiden vastuut oikeudenmukaisella tavalla.

Suuri osa aktivisteista ei pyri täsmällisiin ohjeisiin nopeasti muuttuvalla kentällä, sillä datakysymyksissä poliittinen määrittelykamppailu on kiivasta, eivätkä sitovat normit, kuten EU:n tietosuoja-asetus, kata kuin osan ilmiöstä¹⁰⁵. Data-aktivistit peräänkuuluttavat keskustelua, joka on tarpeeksi kunnianhimoista tarttuakseen myös globaalin sääntelyn puutteisiin ihmisoikeus- ja tasa-arvonäkökulmista. Tasa-arvokysymyksistä on olemassa paljon oivaltavaa ja perusteellista tutkimusta eri tieteenaloilla, mutta tekoälyyn liittyen vielä valitettavan vähän. Viime aikoina on kuitenkin enenevässä määrin ollut nähtävissä tutkijoiden ja aktivistien aloitteita, joissa näkemyksiä kohdennetaan suoraan teknologiayrityksille tasa-arvon ja syrjimättömyyden kannalta keskeisistä kysymyksistä¹⁰⁶. Lisäksi aktivistit tuottavat vaihtoehtoisia teknologioita herättämään keskustelua tärkeistä asioista. Yhtenä esimerkkinä voi mainita feministisen chatbotin nimeltään ”ladymouth”, joka vastaa keskustelupalstan trolleille, eli vihamielisiä tai tahallaan häiriötä aiheuttaville henkilöille. Automaattiset vastaukset lainaavat kuuluisia feministisiä kirjailijoita¹⁰⁷.

MyData OmaData eli MyData¹⁰⁸ on moninainen, digitaalisen ajan ihmis-oikeuksiin, keskittyvä liike, joka on valjastanut aktivisminsa erityisesti edistämään yksilön mahdollisuuksia hallita omien henkilötietojensa hyödyntämistä. Se kamppailee yksityisyyden katoamisen ja datan monopolisoinnin uhkia vastaan. Nykyisessä datataloudessa henkilödatan ostaminen ja myyminen on eräs merkittävimmistä liiketoimintaa muokkaavista voimista. OmaData-liikkeen tärkeimmät periaatteet on kirjattu MyData-julistukseen¹⁰⁹. Suomessa henkilötietojen hallinnan ja käsittelyn periaatteisiin on kiinnitetty runsaasti huomiota: liikenne- ja viestintäministeriö julkaisi jo vuonna 2014 aiheesta selvityksen¹¹⁰ ja aiheeseen liittyviä hankkeita on suunniteltu yhteistyössä julkisen ja yksityisen sektorin sekä useiden järjestöjen ja data-aktivistien kanssa.

¹⁰⁵ Puukko, 2021, s. 149

¹⁰⁶ Puukko, 2023, s. 150

¹⁰⁷ Ciston, 2019

¹⁰⁸ <https://mydata.org/about/>

¹⁰⁹ <https://mydata.org/participate/declaration/>

¹¹⁰ Poikola et al., 2018 (viittaus päivitettyyn versioon – alkuperäinen julkaisu, 2014)

5. Johtopäätökset

Tässä selvityksessä on kuvattu tekoälyn ja sukupuolten tasa-arvon yhtymäkohtia ja nostettu esiin keskeisiä näkökulmia siitä, miten tekoälyjärjestelmät ja niiden kehittäminen sekä käyttö voivat vaikuttaa sukupuolten tasa-arvoon sitä joko heikentäen tai vahvistaen. Selvityksessä todetaan, että usein tekoälyjärjestelmät ylläpitävät ja vahvistavat yhteiskunnallisia ennakkoluuloja ja syrjiviä käytäntöjä. Tekoälyjärjestelmien käyttö on myös tuottanut uusia vahingollisia ilmiöitä, kuten digitaalisen häirinnän ja vainon uusia muotoja. Ne mahdollistavat myös ihmisten nopeamman profiloinnin, ja yksilöiden tunnistamisen seurauksena tekoälyjärjestelmät voivat myös tuottaa päätelmiä, jotka syrjivät sukupuolen perusteella tai ylläpitävät sukupuolistereotyyppioita.

Teknologian nopea ja ennakoimaton kehitys liittyy myös laajoihin geopolitiisiin kysymyksiin. Suurvallat ja suuret teknologiayritykset kilpailevat markkinaosuuksista ja teknologisesta johtajuudesta sekä tehokkaista tavoista tuottaa ja hyödyntää tietoa ihmisistä. Kaikki toimijat eivät noudata kansainvälisiä eettisiä ohjeistuksia, ja jotkut käyttävät tekoälyä tietoisesti moraalisesti kyseenalaisiin tarkoituksiin. Euroopan unionin tekoälyasetuksella pyritään suojaamaan eurooppalaisia suurimmilta väärinkäytöksiltä. Koska asetus koskee lähinnä tekoälyjärjestelmien markkinoille tuomista, on tärkeää ymmärtää, että sukupuolten tasa-arvoa tulee edistää ja tarkastella myös muun lainsäädännön ja hyvien käytäntöjen avulla. Kansallinen ja kansainvälinen perus- ja ihmisoikeuksia koskeva sääntely velvoittaa torjumaan syrjintää ja edistämään sukupuolten tasa-arvoa. Nopeasta ja ennakoimattomasta kehityksestä huolimatta Suomessa on hyvä mahdollisuus kehittää ja käyttää tekoälyä vastuullisesti huomioiden tasa-arvon ja syrjimättömyyden. Tämä vaatii kansallista keskustelua ja linjauksia siitä, kuinka tekoälyjärjestelmien suunnittelu ja arviointi tapahtuu käytännössä perusoikeudet ja sukupuolten tasa-arvo huomioon ottaen. Niin lainsäädännössä kuin tekoälyn eettisissä ohjeistuksissa viitataan yhdenvertaisuuteen ja tasa-arvoon, mutta vain harvoin esitetään konkreettisia tai käytännön toimintaa ohjaavia ratkaisuja ja toimenpiteitä siihen, miten sukupuolten tasa-arvoa voidaan edistää teknologian avulla ja mitä riskien tunnistaminen käytännössä vaatii. Sen tueksi tarvitaan lisää monitieteellistä tutkimusta ja aktiivista poliittista ja kansalaiskeskustelua.

Tekoälyjärjestelmien ongelmat ja rajoitteet eivät ole vain teknisiä ongelmia, joihin on olemassa teknisiä ratkaisuja. Sukupuolten tasa-arvon edistämiseksi tarvitaan laajempaa yhteiskunnallista muutosta ja monipuolisia lähestymistapoja. Koska tekoäly voi vahvistaa ja toistaa olemassa olevia syrjiviä rakenteita ja käytäntöjä, syrjinnän ratkaiseminen vaatii myös yhteiskunnallisia muutoksia ja syrjinnän yhteiskunnallisten mekanismien tunnistamista ja korjaamista. Tekoälysovellukset voivat parhaimmillaan auttaa tunnistamaan ja ratkaisemaan ongelmia, mutta käyttöön liittyvät riskit on tunnistettava ja niitä hallittava.

Selvityksen keskeiset viestit tiivistetysti

- Tekoälyjärjestelmien suunnittelun ja päätelmien pohjana olevan datan takana on monenlaista ihmisten tekemää luokittelua ja datatyötä.
- Tekoälyjärjestelmät voivat ylläpitää ja vahvistavaa yhteiskunnassa vallalla olevia syrjiviä käytäntöjä.
- Tekoälyn ja algoritmisten järjestelmien käyttö on tuottanut uusia syrjinnän muotoja, kuten digitaalinen vainoaminen. Teknologia mahdollistaa myös nopeamman profiloinnin ja yksilöiden tunnistamisen.
- Tekoälyjärjestelmät voivat tuottaa tuloksia, ehdotuksia tai päätelmiä, jotka syrjivät sukupuolen perusteella tai ylläpitävät sukupuolistereotyyppioita. Tämä voi johtua esimerkiksi puuttuvasta tai vinoutuneesta datasta tai siitä, että tekoälyjärjestelmien kehittäjät eivät tunne sovelluksen kontekstia ja siihen liittyviä ongelmia.
- Tekoälyjärjestelmät ovat käytössä monilla sektoreilla ja voivat aiheuttaa sukupuoleen perustuvaa syrjintää monilla eri aloilla kuten mainonnassa, työelämässä, terveydenhuollossa tai lainvalvonnassa.
- Tekoälyn vaikutuksia työelämään, tasa-arvoon ja syrjintään on monin tavoin vaikea ennakoida. Se missä tekoälyä käytetään, on kuitenkin ihmisten päätösten seurausta. Hyvällä suunnittelulla ja työntekijöiden koulutuksella ja osallistamisella kehittämistyöhön voidaan vähentää ei-toivottuja vaikutuksia.
- Tekoälyä voidaan käyttää tunnistamaan syrjiviä rakenteita ja tuottamaan tasa-arvoa.
- Monet kansainväliset ohjeistukset ja säädökset painottavat tai edellyttävät, että tekoälyjärjestelmien kehittämisessä ja käyttöönotossa tulee ottaa huomioon ihmisten perusoikeudet, kuten sukupuolten tasa-arvo. EU:n tekoälyasetuksessa tällainen vaikutustenarviointi on pakollista suurten riskien tekoälyjärjestelmissä, mutta asetuksessa suositellaan, että arviointia tehtäisiin kaikissa tekoälyjärjestelmissä.
- Tekoälyn haitallisia vaikutuksia voi välttää muun muassa huolellisella ja monipuolisella vaikutusten arvioinnilla, käyttökontekstin tuntemisella sekä inklusiivisella kehittämisellä, jossa otetaan huomioon ihmisten erilaiset taustat ja esimerkiksi sukupuoli. Haitallisia vaikutuksia voi vähentää vaikuttamalla yhteiskunnan rakenteisiin ja käytäntöihin esimerkiksi lisäämällä ymmärrystä perus- ja ihmisoikeuksista ja syrjinnän ehkäisystä tekoälyn kehittämisessä ja soveltamisessa.

Lähteet

- Adam, A. 1998. Artificial Knowing: Gender and the Thinking Machine. Routledge. <https://doi.org/10.4324/9780203005057>
- AI Act, 2024. AI Act. European Commission. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
- Alastalo M. & Homanen R. 2015. Hyvinvointivaltion rajankäyntiä maistraatissa: Ulkomaalaisten rekisteröintikäytännöt erilaisten statusten ja valtiollisen tiedon lähteenä. Yhteiskuntapolitiikka 80(2) 147–159.
- Alastalo, M. & Lehto, I. (hyväksytty) Gendered data work practices and the struggle for recognition. A case study of a Finnish public sector organisation building data-driven healthcare and social services. Scandinavian Journal of Information systems.
- Alastalo, M. 2009. Viranomaistiedosta tilastoksi: rekisteriperusteisen tilastojärjestelmän muodostaminen Suomessa. Sosiologia, 46(3). <https://journal.fi/sosiologia/article/view/123881>
- Amnesty International. 2024. Coded Injustice: Surveillance and Discrimination in Denmark's Automated Welfare State. Amnesty International. EUR 18/8709/2024. <https://www.amnesty.org/en/documents/eur18/8709/2024/en/>
- Balch, O. 2020. AI and me: friendship chatbots are on the rise, but is there a gendered design flaw? Guardian 7.5.2020. <https://www.theguardian.com/careers/2020/may/07/ai-and-me-friendship-chatbots-are-on-the-rise-but-is-there-a-gendered-design-flaw>
- Bartoletti, P. I. & Xenidis, R. 2023. Study on the impact of artificial intelligence systems, their potential for promoting equality, including gender equality, and the risks they may cause in relation to non-discrimination. Gender Equality Commission (GEC) and the Steering Committee on Anti-discrimination, Diversity and Inclusion (CDADI). <https://edoc.coe.int/en/artificial-intelligence/11649-study-on-the-impact-of-artificial-intelligence-systems-their-potential-for-promoting-equality-including-gender-equality-and-the-risks-they-may-cause-in-relation-to-non-discrimination.html>
- Bartoletti, P.I. 2020. An Artificial Revolution: On Power, Politics & AI. Indigo Press.
- Benjamin, R. 2019. Race After Technology: Abolitionist Tools for the New Jim Code. Polity Press.
- Bucher, T. 2016. Neither black nor box: Ways of knowing algorithms. Teoksessa Innovative methods in media and communication research, toim. S. Kubitschko & A. Kaun 81–98. Palgrave Macmillan.
- China Law Translate. 2015. Establishment of the Social Credit System. <https://www.chinalawtranslate.com/en/socialcreditsystem/>
- Cifor, M., Garcia, P., Cowan, T.L., Rault, J., Sutherland, T., Chan, A., Rode, J., Hoffmann, A., Salehi, & N., Nakamura, L. 2019. Feminist Data Manifest–No. <https://www.manifestno.com/>
- D'Ignazio, C. & Klein, L. F. 2020. Data Feminism. The MIT Press. <https://doi.org/10.7551/mitpress/11805.001.0001>
- Datta, A., Tschantz, M. C. & Datta, A. 2015. Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. Proceedings on Privacy Enhancing Technologies, 2015(1), 92–112. <https://doi.org/10.1515/popets-2015-0007>
- Dencik, L. 2020. Advancing Data Justice. Teoksessa The Media Manifesto, toim. N. Fenton, D. Freedman, J. Schlosberg & L. Dencik, 48–74. Polity.
- EIGE. 2021. Artificial intelligence, platform work and gender equality. Publications Office of the European Union. https://eige.europa.eu/publications-resources/publications/artificial-intelligence-platform-work-and-gender-equality?language_content_entity=en
- Eubanks, V. 2018. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.
- Euroopan parlamentti. 2021. EUROOPAN PARLAMENTIN JA NEUVOSTON ASETUS tekoälyä koskevista yhdenmukaistetuista säännöistä ja asetusten (EY) N:o 300/2008, (EU) N:o 167/2013, (EU) N:o 168/2013, (EU) 2018/858, (EU) 2018/1139 ja (EU) 2019/2144 sekä direktiivien 2014/90/EU, (EU) 2016/797 ja (EU) 2020/1828 muuttamisesta (tekoälysäädös). <https://data.consilium.europa.eu/doc/document/PE-24-2024-INIT/fi/pdf>

- Fulgu, R. A. & Capraro, V. 2024. Surprising gender biases in GPT. *Computers in Human Behavior Reports*, 16, 100533. <https://doi.org/10.1016/j.chbr.2024.100533>
- Gangadharan, S. & Niklas, J. 2019. Decentering Technology in Discourse on Discrimination. *Information, Communication & Society* 22(7): 882–899. <https://doi.org/10.1080/1369118X.2019.1593484>
- Garcia, M. 2016. Racist in the Machine: The Disturbing Implications of Algorithmic Bias, *World Policy Journal* 33 (2016): 111–117.
- Gillespie, N., Lockey, S., Curtis, C., Pool, J. & Akbari, A. 2023. Trust in Artificial Intelligence: A global study. The University of Queensland; KPMG Australia. <https://doi.org/10.14264/00d3c94>
- GREVIO. 2021. Ensimmäinen yleinen suositus naisiin kohdistuvan väkivallan digitaalisesta ulottuvuudesta. Euroopan neuvosto. <https://rm.coe.int/general-recommendation-no-1-on-the-digital-dimension-of-violence-again/1680ad4d0f>
- Hand, D. J. 2020. *Dark data: Why what you don't know matters*. Princeton University Press.
- Harriger, J. A., Evans, J. A., Thompson, J. K. & Tylka, T. L. 2022. The dangers of the rabbit hole: Reflections on social media as a portal into a distorted world of edited bodies and eating disorder risk and the role of algorithms. *Body Image*, 41, 292–297. <https://doi.org/10.1016/j.bodyim.2022.03.007>
- Helsingin Sanomat. 2020. Kiina pisteyttää yksityishenkilöiden ja pian myös yritysten teot. Pääkirjoitus. <https://www.hs.fi/paakirjoitukset/art-2000006362476.html>
- Hewlett, S. A., Marshall, M. & Sherbin, L. 2013. How Diversity Can Drive Innovation. *Harvard Business Review*. December 2013. <https://hbr.org/2013/12/how-diversity-can-drive-innovation>
- Hoeyer, K. 2023. *Data Paradoxes: The Politics of Intensified Data Sourcing in Contemporary Healthcare*. The MIT Press. <https://doi.org/10.7551/mitpress/14926.001.0001>
- Kauhanen, A. & Rouvinen, P. 2024. Generatiivisen tekoälyn vaikutukset eivät ainakaan vielä näy työmarkkinoilla: Palkat eniten altistuneissa ammateissa pikemminkin nousseet. ETLA Muistio nro 143. <https://pub.etla.fi/ETLA-Muistio-Brief-143.pdf>
- Kinnunen, M. 2022. Luokiteltu sukupuoli. Vastapaino.
- Kuntaliitto. 2024. Kuntien digitalisaatiokartoitus 2024. <https://www.kuntaliitto.fi/julkaisut/digitalisaatiokartoitus-2024>
- Lanier, J. 2020. AI is An Ideology, Not A Technology. *Wired* 15.3.2020 <https://www.wired.com/story/opinion-ai-is-an-ideology-not-a-technology/>
- Larsson, S., White, J. & Ingram Bogusz, C. 2024. The Artificial Recruiter: Risks of Discrimination in Employers' Use of AI and Automated Decision-Making. *Social Inclusion*, 12, 7471. <https://doi.org/10.17645/si.7471>
- Lego, V. D. 2023. Uncovering the gender health data gap. *Cadernos de Saúde Pública*, 39(7), e00065423. <https://doi.org/10.1590/0102-311xen065423>
- Lehtiniemi, T. 2023. Contextual social valences for artificial intelligence: anticipation that matters in social work. *Information, Communication & Society*. 27(6), 1110–1125. <https://doi.org/10.1080/1369118x.2023.2234987>
- Lindroos-Hovinheimo, S. 2024. Lyhyt johdatus unionin tekoälyasetukseen. Teoksessa *Puheenvuoroja sananvapaudesta: Viestintäoikeuden vuosikirja*, toim. P. Korpisaari. 2023, 67–76). (FORUM IURIS). Forum Iuris, Helsingin yliopisto Oikeustieteellinen tiedekunta. <https://www.edilex.fi/viestintaoikeuden-vuosikirja/1000880004.pdf>
- Lovejoy, J. 2018. Fair Is Not the Default: Why building inclusive tech takes more than good intentions. *Google Design*. <https://design.google/library/fair-not-default>
- Milan, S. & Emilio Treré. 2019. Big Data from the South(s): Beyond Data Universalism. *Television & New Media* 20(4): 319–335. <https://doi.org/10.1177/1527476419837739>

Mogensen, C. & Helling Rand. 2020. The angry internet: A threat to gender equality, democracy & well-being. Centre for Digital Youth Care. https://cfdp.dk/wp-content/uploads/2020/11/CFDP_the_angry_internet_ISSUE.pdf

Morsello, B. & Bonomo, C. 2024. Humanizing or feminizing intelligent personal assistants? Exploring the gender representation of Siri, Cortana, and Alexa in the Italian public sphere. *Journal of Gender Studies*, 1–19. <https://doi.org/10.1080/09589236.2024.2431589>

Musizvingoza. R. 2024. Bridging the Gender Data Gap: Harnessing Synthetic Data for Inclusive AI. *UNU Macau*. <https://unu.edu/macau/blog-post/bridging-gender-data-gap-harnessing-synthetic-data-inclusive-ai>

OECD. 2024. Harnessing the green and digital transitions for gender equality: Policy insights from the 2024 OECD Forum on Gender Equality. *OECD Public Governance Policy Papers*. October 2024:61. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/10/harnessing-the-green-and-digital-transitions-for-gender-equality_c9a0091a/860d0901-en.pdf

Ojanen, A., Sahlgren, O., Vaiste, J., Björk, A., Mikkonen, J., Kimppa, K., Laitinen, A. & Oljakka, N. 2022. Algoritminen syrjintä ja yhdenvertaisuuden edistäminen : Arviointikehikko syrjimättömälle tekoälylle. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 2022:54. <http://urn.fi/URN:ISBN:978-952-383-404-0>

Orwat, C. 2024. Algorithmic Discrimination from the Perspective of Human Dignity. *Social Inclusion*, 12, 7160. <https://doi.org/10.17645/si.7160>

Poikola, A., Kuikkaniemi, K., Kuittinen, O., Honko, H. & Knuutila, A. 2018. Johdatus ihmiskeskeiseen henkilötiedon hyödyntämiseen. Liikenne- ja viestintäministeriön julkaisusarja: Muut julkaisut. 3/2018. <http://urn.fi/URN:ISBN:978-952-243-554-5>

Puukko, O. 2021. Kohti dataoikeudenmukaisuutta – ehdotuksia tutkijoilta ja liikkeiltä. *Media & viestintä: kulttuurin ja yhteiskunnan tutkimuksen lehti*, 44 (2) 145–153. <https://doi.org/10.23983/mv.109865>

Ruckenstein, M. 2022a. Algoritmisten systeemien murtumat ja uudet alat (REPAIR):

Tilannekuvaraportti 2022. Suomen Akatemia. <https://www.repair-research.fi/julkaisut/repair-hankkeen-tilanne>

Ruckenstein, M. & Trifuljesko, S. 2022b. Algorithmic futures and the unsettled sense of care. Teoksessa *An Anthropology of Futures and Technologies*, toim. D. Lanzeni, K. Waltrip, S. Pink, & R. C. Smith, 79–92. Taylor and Francis Inc. <https://doi.org/10.4324/9781003084471-6>

Ruckenstein, M. 2023. *The Feel of Algorithms*. University of California Press.

Rusanen, A.–M. 2021. Algoritmien aakkoset. Teoksessa *Älykäs huominen – miten tekoäly ja digitalisaatio muuttavat maailmaa?* Gaudeamus. Salganik, Matthew J et al. 2020. Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences of the United States of America* vol. 117,15 (2020): 8398–8403. doi:10.1073/pnas.1915006117

Salokannel, M. & Ignatius, E. 2023. Lainsäädännölliset ja eettiset reunaehdot ihmisoikeusperustaiselle tiedonkeruulle: Esiselvitys. Oikeusministeriön julkaisuja, Selvityksiä ja ohjeita 2023:35. <http://urn.fi/URN:ISBN:978-952-400-807-5>

Similä, V. 2024. Näin puhui Andrew Tate. *Helsingin Sanomat*. 17.2.2024. <https://www.hs.fi/kirjeenvaihtajat/art-2000010842196.html>

Teknologiaeollisuus. 2024. Tekniikan alan korkeakoulutuksen hakijamäärä kasvaa – naisten osuus saisi kehittyä reippaammin. <https://teknologiaeollisuus.fi/tekniikan-alan-korkeakoulutuksen-hakijamaara-kasvaa-naisten-osuus-saisi-kehittyä-reippaammin/>

Tekoälyä käsittelevä korkean tason asiantuntijaryhmä (AI HLEG). 2019. Luotettavaa tekoälyä koskevat eettiset ohjeet. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60426

Toivonen, H. 2023. Mitä tekoäly on? 100 kysymystä ja vastausta. Teos.

UNESCO, OECD, IDB. 2022. The Effects of AI on the Working Lives of Women. OECD. https://www.oecd.org/en/publications/2022/03/the-effects-of-ai-on-the-working-lives-of-women_1b627535.html

UNESCO. 2022. Recommendation on the Ethics of Artificial Intelligence. <https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence>

van Gerven, M. 2024. Syrjivät algoritmit nakertavat luottamusta Pohjolan hyvinvointivaltioissa – on aika unohtaa keppi ja panostaa tukeen. Repair Research. <https://www.repair-research.fi/blogi/syrjivt-algoritmit-nakertavat-luottamusta-pohjolan-hyvinvointivaltioissa-nbspon-aika-unohtaa-keppi-ja-panostaa-tukeen>

Viljanen, M. 2023. Menikö juna jo? Tekoälyn sääntelemisen mahdollisuuksista. Lakimies. 121(7-8), 1204-1231. <https://journal.fi/lakimies/article/view/136265>

Woods, H. S. 2018. Asking more of Siri and Alexa: feminine persona in service of surveillance capitalism. Critical Studies in Media Communication, 35(4), 334-349. <https://doi.org/10.1080/15295036.2018.1488082>

World Economic Forum. 2023. Global Gender Gap Report. https://www3.weforum.org/docs/WEF_GGGR_2023.pdf

Yhdenvertaisuus- ja tasa-arvolautakunta. 2018. Täysistunto (ään.) Luottokelpoisuuden arviointimenettely, toimivalta, välitön moniperusteinen syrjintä, sukupuoli, kieli, ikä, asuinpaikka, taloudelliset syyt, uhkasakko Diaarinumero: 217/2017 Antopäivä: 21.3.2018

Yhdenvertaisuusvaltuutettu. 2022a. Yhdenvertaisuusvaltuutetun kertomus eduskunnalle 2022. [https://yhdenvertaisuusvaltuutettu.fi/documents/25249352/42720545/Yhdenvertaisuusvaltuutetun%20kertomus%20eduskunnalle%202022%20\(PDF\).pdf](https://yhdenvertaisuusvaltuutettu.fi/documents/25249352/42720545/Yhdenvertaisuusvaltuutetun%20kertomus%20eduskunnalle%202022%20(PDF).pdf)

Yhdenvertaisuusvaltuutettu. 2022b. Yhdenvertaisuusvaltuutetun suositukset hallitusohjelmaan 2023-2027. [https://yhdenvertaisuusvaltuutettu.fi/documents/25249352/143315235/Yhdenvertaisuusvaltuutetun+suositukset+hallitusohjelmaan+\(PDF\).pdf/dd8d4ddf-2ab7-9c25-437d-c6619da62712/Yhdenvertaisuusvaltuutetun+suositukset+hallitusohjelmaan+\(PDF\).pdf?t=1681804769918](https://yhdenvertaisuusvaltuutettu.fi/documents/25249352/143315235/Yhdenvertaisuusvaltuutetun+suositukset+hallitusohjelmaan+(PDF).pdf/dd8d4ddf-2ab7-9c25-437d-c6619da62712/Yhdenvertaisuusvaltuutetun+suositukset+hallitusohjelmaan+(PDF).pdf?t=1681804769918)

Ylikoski, P. 2024. Tekoäly ja asiantuntijatyön muutos. Akava Works. [https://akavaworks.fi/julkaisut/artikkelit/nakokulmia-tekoalyyn-osa-5-petri-ylikoski-tekoaly-ja-asiantuntijatyon-muostos/](https://akavaworks.fi/julkaisut/artikkelit/nakokulmia-tekoalyyn-osa-5-petri-ylikoski-tekoaly-ja-asiantuntijatyon-muutos/)

Ylikoski, P. & Snell, K. 2021. Datalukutaitoon tarvitaan monipuolista ymmärrystä. <https://www.datalit.fi/datalukutaitoon-tarvitaan-monipuolista-ymmarrysta/>



Tane

Tasa-arvoasiain
neuvottelukunta