

Syrjivyyden tunnistettavuus luonnollisia henkilöitä koskevan algoritmisen päätöksenteon yhteydessä

Jere Lehtioksa

Helsingin yliopisto

Oikeustieteellinen tiedekunta

Hallinto-oikeuden projektin OTM-tutkielma

Ohjaajat: Ida Koivisto – Riikka Koulu

Huhtikuu 2022

Tiivistelmä

Tiedekunta: Oikeustieteellinen tiedekunta

Koulutusohjelma: Oikeustieteen maisterin koulutusohjelma

Opintosuunta: Hallinto-oikeus

Tekijä: Jere Lehtioksa

Työn nimi: Syrjivyyden tunnistettavuus luonnollisia henkilöitä koskevan algoritmisen päätöksenteon yhteydessä

Työn laji: OTM-tutkielma

Kuukausi ja vuosi: Huhtikuu 2022

Sivumäärä: 76 + XIII

Avainsanat: tietosuoja, hallinto-oikeus, julkishallinto, automaattinen päätöksenteko, koneoppiminen, läpinäkyvyys, avoimuus, syrjintä, yhdenvertaisuus, tasa-arvo, oikeusturva

Ohjaaja tai ohjaajat: Ida Koivisto, Riikka Koulu

Säilytyspaikka: Helsingin yliopiston kirjasto

Muita tietoja:

Tiivistelmä:

Tutkielmassa analysoin algoritmisen päätöksenteon mahdollisesti sisältävän piilevän syrjivyyden tunnistamisen haasteita päätöksenteon kohteena olevan henkilön näkökulmasta tarkasteltuna. Tutkielmassa arvioin sitä, millä tavoin algoritmisessa päätöksenteossa hyödynnettävissä koneoppimiseen perustuvissa tekoälyjärjestelmissä voi ilmetä syrjivyyttä. Tutkimuksen keskeinen kysymys on se, onko läpinäkyvyyden ja avoimuuden kautta ylipäänsä mahdollista luoda riittävä ymmärrys syrjivyydestä päätöksenteon kohteena olevalle henkilölle, jotta tämän olisi mahdollista reagoida asiaan.

Tutkielmassa käytetään useita eri oikeustieteellisessä tutkimuksessa sovellettavia metodologioita, joista keskeisin on tietosuojaoikeuden ja hallinto-oikeuden osa-alueiden systematisointiin perustuva lainopillinen menetelmä. Tutkielma sisältää myös analyttisen oikeustieteen menetelmän mukaista tutkimusta. Tutkielma sisältää elementtejä myös oikeus- ja yhteiskuntatieteellisestä tutkimuksesta, sillä tutkielmassa analysoidaan oikeutta sen yhteiskunnallisessa kontekstissa.

Tutkielman olennainen havainto on se, että algoritmisen päätöksenteon syrjivyyks voi saada alkunsa eri tavoin päätöksenteossa: kyseeseen saattaa tulla yhteiskunnassa vallitsevien epäkohtien tahaton hyödyntäminen tekoälyjärjestelmissä tai tarkoituksellinen toiminta, jossa päätöksenteossa hyödynnetään sellaisia valintoja, jotka johtavat syrjintään. Tästä seuraa, että syrjivyyden tunnistamista tulee lähestyä tapauskohtaisesti. Ongelmallisemmaksi muodostuu korrelaatio-suhteisiin perustuvan päätöksenteon luoma niin sanottu piilevä syrjintä, jota ei käytännössä ole mahdollista tunnistaa järjestelmän mustan laatikon sisältä. Tällöin hallinnon algoritmisen päätöksenteon syrjivyyden tunnistamista koskevan keskustelun osalta erityishuomion tulisi kiinnittyä tällaisten päätöksentekojärjestelmien riskien asianmukaiseen hallintaan.

Sisällysluettelo

LYHENNELUETTELO	III
LÄHTEET	IV
1. JOHDANTO	1
1.1. Tutkimuksen aihe ja tutkimuskysymykset.....	1
1.2. Tutkimusmetodi ja lähteet	7
1.3. Keskeisiä rajauksia	11
1.4. Tutkielman rakenne	12
2. ALGORITMISEN PÄÄTÖKSENTEON TEKNINEN TOTEUTUS JA SYRJIVYYDEN ILMENEMINEN TEKOÄLYJÄRJESTELMISSÄ	14
2.1. Algoritmisen päätöksenteon tekninen toteutus.....	14
2.1.1. Koneoppiminen.....	14
2.2. Profiloinnin ja automaattisen päätöksenteon määritelmä EU:n tietosuojalainsäädännössä	16
2.2.1. Tietosuoja-asetuksen mukainen automaattinen päätöksenteko	16
2.2.2. Profilointi	17
2.3. Hallintopäätökset ja tosiasiallinen hallintotoiminta	19
2.3.1. Hallintopäätösten ja tosiasiallisen hallintotoiminnan automatisoinnin nykytila.....	20
2.4. Algoritmisten päätöksentekojärjestelmien syrjivyyden.....	22
2.4.1. Johdanto algoritmisten päätöksentekojärjestelmien syrjivyyteen.....	22
2.4.2. Algoritmisen päätöksenteon syrjivyyden seuraukset	26
3. TAUSTALLA VAIKUTTAVA LAINSÄÄDÄNTÖ	28
3.1. Perusoikeudet.....	28
3.1.1. Oikeus yhdenvertaiseen kohteluun.....	29
3.1.2. Oikeus yksityiselämän suojaan	31
3.2. Syrjintää koskeva kansallinen lainsäädäntö	32
3.3. Tietosuoja-asetus ja automaattinen päätöksenteko	35
4. ALGORITMISEN PÄÄTÖKSENTEON LÄPINÄKYVYYS JA YMMÄRRETTÄVYYS	39
4.1. Läpinäkyvyys laajasti ymmärrettynä	39
4.2. Henkilötietojen käsittelyn läpinäkyvyys	42
4.2.1. Läpinäkyvyys tietosuoja-asetuksessa.....	44
4.2.2. Läpinäkyvyyden toteutustapa tietosuoja-asetuksessa	49
4.3. Tietosuoja-asetuksen mukainen automaattisen päätöksenteon selitysoikeus	51
4.4. Päätösten perusteluvollisuus.....	54
4.4.1. Perusteluvollisuuden toteuttaminen algoritmisen päätöksenteon osalta	55

4.5.	Algoritmista päätöksentekoa koskevien tietojen toimittaminen suhteessa päätöksenteon elinkaareen	58
4.5.1.	Ennen päätöksen tekemistä rekisteröidylle annettavat tiedot.....	59
4.5.2.	Päätöksenteon jälkeen annettavat tiedot.....	62
4.5.3.	Johtopäätökset algoritmista päätöksentekoa koskevien tietojen toimittamisesta suhteessa algoritmisen päätöksenteon elinkaareen	64
5.	HAASTEET ALGORITMISEN PÄÄTÖKSENTEON SYRJIVYYDEN TUNNISTAMISESSA	66
5.1.	Rekisteröidyn kyvyttömyys ymmärtää algoritmista päätöksentekoa koskevaa tietoa	66
5.1.1.	Päätöksentekojärjestelmän kehitysvaiheen tiedot.....	68
5.1.2.	Järjestelmien soveltamat korrelaatiot ja piilevä syrjivyyys	70
5.1.3.	Johtopäätökset rekisteröidyn kyvyttömyydestä ymmärtää algoritmista päätöksentekoa koskevaa tietoa	71
6.	JOHTOPÄÄTÖKSET	74

LYHENNELUETTELO

EU Euroopan unioni

HE hallituksen esitys

PeVL perustuslakivaliokunnan lausunto

SEU Euroopan unionista tehty sopimus

LÄHTEET

Kirjallisuus ja artikkelit

Ananny, Mike – Crawford, Kate, Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New media & society* 2018, 20(3) 2018, s. 973–989

Andreeva, Galina – Matuszyk, Anna, The law of equal opportunities or unintended consequences?: The effect of unisex risk assessment in consumer credit. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182. 10.1111/rssa.12494. 2019, s. 1–25

Anrig, Bernhard – Browne, Will – Gasson, Mark: *The Role of Algorithms in Profiling*, s. 65–87 teoksessa Hildebrandt, Mireille – Gutwirth, Serge: *Profiling the European Citizen: Cross-Disciplinary Perspectives*. Springer, New York 2008

Anttila, Outi – Ojanen, Tuomas: *Yhdenvertaisuuslaki kommentein*. Alma Talent Helsinki 2019, sähköinen teos (sähköinen kirja on sisällöltään sama kuin vuonna 2019 ilmestynyt samanniminen kirja)

Barocas, Solon – Selbst, Andrew D., *Big Data's Disparate Impact*. *California Law Review* 104(3) 2016, s. 671–732

Barrett, Lindsey, *Reasonably Suspicious Algorithms: Predictive Policing at the United States Border*. *N.Y.U. Rev. L. & Soc. Change* 41, 2017, s. 327–363, saatavilla SSRN: <<https://ssrn.com/abstract=2863355>> (haettu 21.3.2022)

Brkan, Maja, *Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond*. Paper submitted in view of presentation at the conference ‘Terminator or the Jetsons? The Economics and Policy Implications of Artificial Intelligence’, Technology Policy Institute, Washington 22 February 2018, 2018, s. 1–29, saatavilla SSRN: <<https://ssrn.com/abstract=3124901>> tai <<http://dx.doi.org/10.2139/ssrn.3124901>> (haettu 12.3.2022)

Burrell, Jenna, How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* January–June 2016, s. 1–12

Citron, Danielle Keats – Pasquale, Frank, The Scored Society: Due Process for Automated Predictions. University of Maryland Francis King Carey School of Law, Legal Studies Research Paper No. 2014 – 8, 2014, s. 1–33

Diakopoulos, Nicholas, Accountability in Algorithmic Decision Making. *Communications of the ACM*, 59(2) February 2016, 2016, s. 56–62

Edwards, Lilian – Veale, Michael, Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For. *Duke Law & Technology Review* 16, 2017, s. 18–84, saatavilla SSRN: <<https://ssrn.com/abstract=2972855>> tai <<http://dx.doi.org/10.2139/ssrn.2972855>> (haettu 12.3.2022)

Goldenfein, Jake, Algorithmic Transparency and Decision-Making Accountability: Thoughts for buying machine learning algorithms. In Office of the Victorian Information Commissioner (ed), *Closer to the Machine: Technical, Social, and Legal aspects of AI* (2019), 2019, s. 41–145, saatavilla SSRN: <<https://ssrn.com/abstract=3445873>> (haettu 15.3.2022)

Hakkarainen, Jenni – Koulu, Riikka – Markkanen, Kalle, Läpinäkyvät algoritmit? Lähdekoodin julkisuus ja laillisuuskontrolli hallinnon digitalisaatiossa. *Edilex 2020/18 Referee-artikkeli*, 2020, s. 1–52

Hardt, Moritz – Price, Eric – Srebro, Nathan, Equality of Opportunity in Supervised Learning. arXiv:1610.02413v1 [cs.LG], 2016, s. 1–22, saatavilla: <<https://doi.org/10.48550/arXiv.1610.02413>> (haettu 23.3.2022)

Hildebrandt, Mireille, The Dawn of a Critical Transparency Right for the Profiling Era. *Digital Enlightenment Yearbook 2012*, J. Bus et al. (Eds.) IOS Press, 2012, s. 41–56

Hildebrandt, Mireille: Defining Profiling: A New Type of Knowledge?, s. 17–45 teoksessa Hildebrandt, Mireille – Gutwirth, Serge: Profiling the European Citizen: Cross-Disciplinary Perspectives. Springer, New York 2008

Hirvonen, Ari, Mitkä metodit? Opas oikeustieteen metodologiaan. Yleisen oikeustieteen julkaisuja 17, Helsinki 2011

Hirvonen, Hanne, Automatisoitu päätöksenteko julkisella sektorilla. Oikeus 47(3) 2018, s. 302–310

Janssen, Heleen L., An approach for a fundamental rights impact assessment to automated decision-making. International Data Privacy Law, 10(1) 2020, s. 76–106

Koivisto, Ida, Thinking Inside the Box: the Promise and Boundaries of Transparency in Automated Decision-Making. Academy of European Law, EUI Working Papers, AEL 2020/01, 2020

Korpisaari, Päivi – Pitkänen, Olli – Warmo-Lehtinen, Eija, Uusi tietosuojalainsäädäntö. Alma Talent Oy Helsinki 2018, sähköinen teos

Koulu, Riikka, Digitalisaatio ja algoritmit – oikeustiede hukassa? Lakimies 7–8/2018, 2018, s. 840–867

Kroll, Joshua A. – Huey, Joanna – Barocas, Solon – Felten, Edward W. – Reidenberg, Joel R. – Robinson, David G. – Yu, Harlan, Accountable Algorithms. U. Pa. L. Rev. 165, 2017, s. 633–706 saatavilla: <https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3> (haettu 22.3.2022)

Lepri, Bruno – Oliver, Nuria – Letouzé, Emmanuel – Pentland, Alex – Vinck, Patrick, Fair, Transparent, and Accountable Algorithmic Decision-Making Processes. Philosophy & Technology 31(4), 2018, 611–627, s. 1-18, saatavilla: <https://dspace.mit.edu/bitstream/handle/1721.1/122933/13347_2017_279_ReferencePDF.pdf;jsessionid=699F826C0A43104FFA70B889CB15E393?sequence=2> (haettu 23.3.2022)

Malgieri, Gianclaudio – Comandé, Giovanni, Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation. *International Data Privacy Law* 7(4) 2017, s. 243–265

Minkkinen, Panu, Oikeus- ja yhteiskuntatieteellinen tutkimus – suuntaus, tarkastelutapa, menetelmä? *Lakimies* 7–8/2017, 2017, s. 908–923

Mäenpää, Olli: *Hallinto-oikeus*. Alma Talent Oy, sähköinen teos (sisältö perustuu vuonna 2018 Oikeuden perusteokset -sarjassa julkaistuun *Hallinto-oikeus*-teoksen toiseen, uudistettuun painokseen), päivitetty kesäkuussa 2021

O’Hara, Kieron, Explainable AI and the philosophy and practice of explanation. *Computer law & security review* 39, 105474, 2020, s. 1–7

Pasquale, Frank, *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press 2015, sähköinen teos

Roscher, Ribana – Bohn, Bastian – Duarte, Marco F. – Garcke, Jochen, Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access* 8, 2020 s. 42200-42216, doi: 10.1109/ACCESS.2020.2976199

Selbst, Andrew D. –Barocas, Solon, The Intuitive Appeal of Explainable Machines. *Fordham Law Review* 87, 2018, s. 1085–1139, saatavilla: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3126971> (haettu 7.4.2022)

Soininen, Niko, Tulkinnan pelko (ja kuinka se voitetaan perusteluvollisuudella), *Lakimies* 7–8/2017, 2017, s. 1145–1149

Stohl, Cynthia – Stohl, Michael – Leonardi, Paul M., Managing Opacity: Information Visibility and the Paradox of Transparency in the Digital Age. *International Journal of Communication* 10, 2016, s. 123–137

Van der Hof, Simone – Prins, Corien: Personalisation and its Influence on Identities, Behaviour and Social Values, s. 111–127 teoksessa Hildebrandt, Mireille – Gutwirth, Serge: Profiling the European Citizen: Cross-Disciplinary Perspectives. Springer, New York 2008

Veale, Michael – Binns, Reuben, Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. Big Data & Society, July–December 2017, s. 1–17

Voutilainen, Tomi, Automatisoitu hallintoasian käsittelyprosessi. Edilex referee-artikkeli (versio 1.0), 2008/12, 2008, s. 1–44

Wachter, Sandra – Mittelstadt, Brent – Russell, Chris, Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. Harvard Journal of Law & Technology, 31(2) 2018, s. 841-888

Wachter, Sandra – Mittelstadt, Brent – Floridi, Luciano, Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 2017, s. 76–99

Zarsky, Tal, The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. Science, Technology & Human Values, 41(1), 2016, s. 118-132

Zarsky, Tal Z., Understanding Discrimination in the Scored Society. Washington Law Review, 89(4), 2014, s. 1375-1412, saatavilla: SSRN: <<https://ssrn.com/abstract=2550248>> (haettu 21.3.2022)

Zliobaite, Indre, A survey on measuring indirect discrimination in machine learning. arXiv:1511.00148v1 [cs.CY], 2015, s. 1–21, saatavilla: <<https://doi.org/10.48550/arXiv.1511.00148>> (haettu 21.3.2022)

Virallislähteet

Kansallinen lainvalmisteluaineisto

HE 18/2019 vp, Hallituksen esitys eduskunnalle laiksi henkilötietojen käsittelystä maahanmuuttohallinnossa ja eräksi siihen liittyviksi laeiksi

HE 19/2014 vp, Hallituksen esitys eduskunnalle yhdenvertaisuuslaiksi ja eräksi siihen liittyviksi laeiksi

HE 72/2002 vp, Hallituksen esitys eduskunnalle hallintolaiksi ja laiksi hallintolainkäyttölain muuttamisesta

HE 96/1998 vp, Hallituksen esitys eduskunnalle henkilötietolaiksi ja eräksi siihen liittyviksi laeiksi

Valiokunnan lausunto PeVL 7/2019 vp - HE 18/2019 vp Hallituksen esitys eduskunnalle laiksi henkilötietojen käsittelystä maahanmuuttohallinnossa ja eräksi siihen liittyviksi laeiksi

Valiokunnan lausunto PeVL 14/2018 vp. HE 9/2018 vp Hallituksen esitys eduskunnalle EU:n yleistä tietosuojaa-asetusta täydentäväksi lainsäädännöksi

Perustuslakivaliokunnan lausunto 37/2014 vp, hallituksen esitys eduskunnalle laeiksi perusopetuslain 13 §:n ja lukiolain 9 §:n muuttamisesta HE 136/2014 vp

Kansalliset viranomaisjulkaisut ja lausunnot

Hallinnon automaattista päätöksentekoa koskevaa yleislainsäädäntöä valmisteleva työryhmä, muistio: hallinnon automaattinen päätöksenteko, käyttöalaa ja läpinäkyvyyttä koskevat säännösluonnokset, 31.5.2021, saatavilla: <<https://www.lausuntopalvelu.fi/FI/Proposal/Participation?proposalId=0cca5af7-764a-4f6c-a4a4-251b1e2939f5>> (haettu 7.4.2022)

Oikeuskanslerin lausunto, 19.8.2021. OKV/1698/21/2021

Oikeusministeriö, julkisen hallinnon automaattista päätöksentekoa koskeva lainsäädäntö, työryhmämietintö 2022:7, 2022, saatavilla: <<https://julkaisut.valtioneuvosto.fi/handle/10024/163847>> (haettu 7.4.2022)

Oikeusministeriö, Arviomuistio hallinnon automaattiseen päätöksentekoon liittyvistä yleislainsäädännön sääntelytarpeista, 2020:14, 2020, saatavilla: <<https://julkaisut.valtioneuvosto.fi/handle/10024/162355>> (haettu 7.4.2022)

Valtioneuvosto, Tekoälyn kokonaiskuva ja kansallinen osaamiskartoitus – loppuraportti 4/2019, 2019, saatavilla: <https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161282/4-2019_Tekoalyn%20kokonaiskuva.pdf?sequence=1&isAllowed=y> (haettu 15.3.2022)

Yhdenvertaisuusvaltuutetun lausunto, arviomuistio hallinnon automaattiseen päätöksentekoon liittyvistä yleislainsäädännön sääntelytarpeista VVTDno-2020–620, Asia: VN/3071/2020

Kansainvälinen virallisaineisto

Euroopan komission valkoinen kirja tekoälystä – Eurooppalainen lähestymistapa huippuosaamiseen ja luottamukseen 2020, COM(2020) 65 final

Euroopan parlamentti, tutkimus, Understanding algorithmic decision-making: Opportunities and challenges, 2019, saatavilla: <[https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU\(2019\)624261_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf)> (haettu 10.3.2022)

29 Artiklan mukainen tietosuojatyöryhmä, suuntaviivat automatisoiduista yksittäispäätöksistä ja profiloinnista asetuksen (EU) 2016/679 täytäntöön panemiseksi 17/FI, WP251rev.01, 2017

29 Artiklan mukainen tietosuojatyöryhmä, asetuksen 2016/679 mukaista läpinäkyvyyttä koskevat suuntaviivat 17/FI WP260 rev.01, 2017

Oikeustapaukset

Euroopan unionin tuomioistuin

Julkisasiamies Cruz Villalónin 9. heinäkuuta 2015 asiassa Bara (C-201/14) antama ratkaisuehdotus

Yhdenvertaisuus- ja tasa-arvolautakunta

Yhdenvertaisuus- ja tasa-arvolautakunnan päätös, 21.3.2018, 216/2017

Valtioneuvoston oikeuskansleri

Valtioneuvoston oikeuskansleri, oikeuskanslerin ratkaisu 2021, OKV/131/70/2020
20.4.2021

Muut lähteet

Eduskunnan oikeusasiamies, Verohallinnon automatisoitu päätöksentekomenettely ei täytä perustuslain vaatimuksia - Tiedotteet, 26.11.2019, saatavilla:
<<https://www.oikeusasiamies.fi/fi/-/verohallinnon-automatisoitu-paatoksentekomenettely-ei-tayta-perustuslain-vaatimuksia>> (haettu 11.3.2022)

Eduskunnan oikeusasiamies, oikeusasiamies ja apulaisoikeusasiamiehet, saatavilla:
<<https://www.oikeusasiamies.fi/fi/oikeusasiamies-ja-apulaisoikeusasiamiehet>> (haettu 15.3.2022)

Euroopan komissio, A European approach to artificial intelligence, saatavilla:
<<https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>>
Home>Policies>A European approach to artificial intelligence (haettu 21.3.2022)

Euroopan tietosuojaneuvosto, Tietosuojatyöryhmä WP 29, saatavilla:
<https://edpb.europa.eu/about-edpb/more-about-edpb/article-29-working-party_fi>Home>About EDPB>Tietosuojatyöryhmä WP 29 (haettu 15.3.2022)

Euroopan tietosuojaneuvosto, EDPB Statement : EDPB cooperation on the elaboration of guidelines, 2021, saatavilla: <https://edpb.europa.eu/news/news/2021/edpb-statement-edpb-cooperation-elaboration-guidelines_en>Home>News>EDPB Statement: EDPB cooperation on the elaboration of guidelines (haettu 15.3.2022)

Lexico.com, artificial intelligence, saatavilla: <https://www.lexico.com/definition/artificial_intelligence> (haettu 15.3.2022).

Machinelearningmastery.com, Difference Between Algorithm and Model in Machine Learning, saatavilla: <<https://machinelearningmastery.com/difference-between-algorithm-and-model-in-machine-learning/>> (haettu 15.3.2022)

Machinelearningmastery.com, Machine Learning Algorithms Mini-Course, saatavilla: <<https://machinelearningmastery.com/machine-learning-algorithms-mini-course/>> (haettu 21.3.2022)

Merriam-Webster, neural network, saatavilla: <<https://www.merriam-webster.com/dictionary/neural%20network>> (haettu 16.3.2022)

Oikeusministeriö, Automaattista päätöksentekoa koskevan hallinnon yleislainsäädännön valmistelu, OM021:00/2020, 2020, saatavilla: <<https://oikeusministerio.fi/hanke?tunnus=OM021:00/2020>>

Oikeusministeriö>FI>Hankkeet ja säädösvalmistelu>Hanke (haettu 11.3.2022)

Tietosuojavaltuutetun toimisto, automaattinen päätöksenteko ja profilointi, saatavilla: <<https://tietosuoja.fi/automaattinen-paatoksenteko-profilointi>>

Organisaatiot>Henkilötietojen käsittely>Automaattinen päätöksenteko ja profilointi (haettu 23.3.2022).

Yhdenvertaisuusvaltuutettu, yhdenvertaisuusvaltuutettu, saatavilla: <<https://syryjinta.fi/etusivu>> (haettu 15.3.2022)

Yhdenvertaisuusvaltuutettu, tekoäly ja yhdenvertaisuus, saatavilla:

<<https://syrjinta.fi/tekoaly#syrjintakiellon-valvonta>> Etusivu>Syrjintä ja yhdenvertaisuus>Aiheet>Tekoäly (haettu 21.3.2022)

6Aika, Espoon tekoälykokeilun jatko: syrjäytymisriskin tekijöihin päästy kiinni, merkittäviä kustannussäästöjä jo nyt, saatavilla: <<https://6aika.fi/espooon-tekoalykokeilun-jatko-syrjaytymisriskin-tekijoihin-paasty-kiinni-merkittavia-kustannussaastoja-jo-nyt/>> Espoo>Uutiset>Asiakkuus, 23.10.2019 (haettu 17.3.2022)

1. JOHDANTO

1.1. Tutkimuksen aihe ja tutkimuskysymykset

Algoritmisia päätöksentekojärjestelmiä hyödynnetään lisääntyvässä määrin päätöksenteon yhteydessä yhteiskunnan eri osa-alueilla.¹ Tässä tutkielmassa algoritmisella päätöksentekojärjestelmällä tarkoitetaan päätöksenteon yhteydessä hyödynnettäviä algoritmeihin perustuvia järjestelmiä.² Algoritmiset päätöksentekojärjestelmät on mahdollista jaotella koneoppimiseen perustuviin ja toisaalta sääntöpohjaisiin järjestelmiin.³

Koneoppimiseen perustuva tekoälyjärjestelmä kykenee löytämään suuresta määrästä aineistoa (kuten esimerkiksi hakemukset ja niistä tehdyt hallintopäätökset) hakemuksen ja päätöksen kesken vallitsevia säännönmukaisuuksia. Tämän jälkeen järjestelmä soveltaa havaitsemaansa tilastollista säännönmukaisuutta uusiin hakemuksiin. Koneoppimisjärjestelmien haaste lain soveltamisen kannalta tarkasteltuna on se, että kyseeseen tulee vain todennäköisyyksiin perustuva ennuste oikeudellisen arvion sijasta.⁴

Algoritmisen päätöksenteon käsitteen sisälle tavataan laskea myös sellainen koneoppimiseen pohjautuva automaatio, joka muodostaa ihmisen suorittamaa päätöksentekoa tukevia analyysejä ja suosituksia.⁵ Tällöin päätöksenteon riskejä on mahdollista arvioida laajemmin sisältäen myös tilanteet, joissa lopullisen päätöksen tekee ihminen koneen luoman luokittelun perusteella.⁶ Esimerkki tästä olisi tilanne, jossa tekoälyjärjestelmä luokittelee hakijoita ja tekee tämän perusteella analyysejä ja ratkaisuehdotuksia, jotka ihminen hyväksyy tosiasiallisesti kuitenkin vaikuttamatta merkityksellisellä tavalla lopputulokseen.⁷

¹ Selbst–Barocas 2018, s. 1087.

² Tässä tutkielmassa viitataan myös automaattiseen päätöksentekoon, jolloin kyseeseen tulee nimenomaan tietosuoja-asetuksessa määritelty automaattinen päätöksenteko tai muussa viranomaislähteessä tai virallislähteessä käytetty termi ”automaattinen päätöksenteko”. Algoritmien päätöksentekojärjestelmä on kuitenkin nähdäkseni käsitteenä laajempi, sillä se ei rajoitu tietosuoja-asetuksen automaattisen päätöksenteon määritelmään. Algoritmisen päätöksentekojärjestelmä kuitenkin pitää sisällään tässä tutkielmassa myös tietosuoja-asetuksen mukaisen automaattisen päätöksenteon määritelmän.

³ Euroopan parlamentin tutkimus 2019, s. 3–4. Hakkaraisen ym. mukaan täydellistä erottelua sääntöpohjaiseen ja tekoälyvetoiseen päätösautomaatioon ei ole tarpeen tehdä, sillä on tavallista, että useita teknisiä toteutustapoja yhdistellään järjestelmissä. Ks. Hakkarainen ym. 2020, s. 2.

⁴ Oikeusministeriön arviomuistio 2020, s. 12.

⁵ Euroopan parlamentin tutkimus 2019, s. 3–4.

⁶ Toisaalta on havaittavissa päätöksentekoa, joka on luonteeltaan korostetun mekaanista (esimerkiksi ”päätos” rakennuksen ulko-oven avaamisesta ja sulkemisesta liikkeen tunnistukseen pohjautuen), eikä siten luo samankaltaisia päätöksenteon ymmärrettävyyteen ja syrjivyyteen liittyviä riskejä.

⁷ Tietosuoja-asetuksessa asetetaan reunaehdot automaattisen päätöksenteon hyödyntämiselle. Tietosuoja-asetuksen automaattisen päätöksenteon määritelmä kattaa päätökset, jotka perustuvat pelkästään automaattiseen käsittelyyn. Tällä tarkoitetaan sitä, että päätöksentekoprosessiin ei osallistu ihmistä. Tästä seuraa, että jos automatisoitu prosessi tuottaa rekisteröityä koskevan suosituksen, jota ihminen kuitenkin tarkastelee huomioiden muita tekijöitä lopullista päätöstä

Algoritmisten päätöksentekojärjestelmien kehityksen ovat osaltaan mahdollistaneet teknologian edistysaskeleet kuten datan kerryttämistä ja sen analysointia tukevat teknologiat ja koneoppimisen kehitysaskeleet. Nimenomaan henkilötietojen laaja saatavuus sekä tekoälyjärjestelmien kyvykkyys havaita korrelaatioita ja luoda yhteyksiä asioiden välille ovat mahdollistaneet henkilöiden luonteenpiirteisiin, käyttäytymiseen, kiinnostuksenkohteisiin ja tapoihin liittyvien seikkojen ennustamisen ja analysoinnin ennennäkemättömällä tavalla.⁸

Algoritmisen päätöksenteon tuomat hyödyt ovat kiistattomia ulottuen sekä yhteiskunta-, että yksilötasolle.⁹ Yksilön kannalta algoritmisen päätöksenteon hyödyt liittyvät myös eri toimintojen tehokkuuden lisääntymiseen ja toisaalta resurssien säästämiseen, joista voi ideaalitulanteissa seurata myönteisiä vaikutuksia yksilöille.¹⁰

Myös julkisella sektorilla algoritmisen päätöksenteko voi edistää ja tehostaa asioiden hoitamista. Tästä konkreettisena esimerkkinä on hallinnon organisaatioiden päätöksentekoprosessin tukeminen ja tehostaminen, jolloin päätökset on mahdollista suorittaa tehokkaammin algoritmisen päätöksenteon avulla verrattuna perinteiseen ihmisjohtoiseen päätöksentekoon.¹¹

Etenkin julkisen sektorin algoritmisen päätöksenteon yhteydessä on mielestäni tarpeen kiinnittää huomioita yksittäisen päätöksenteon kohteena olevan henkilön asemaan.¹² Luonnollisen henkilön asema suhteessa algoritmiseen päätöksentekoon onkin noussut julkiseen keskusteluun viime aikoina esimerkiksi Verohallintoa koskevassa eduskunnan apulaisoikeusasiamiehen päätöksessä, jonka mukaan hyvän hallinnon perusteiden mukaisesta luottamuksensuojasta seuraa, että asiakkaille tulee tiedottaa avoimesti, milloin heidän asiansa on käsitelty automaattisessa päätöksentekomenettelyssä. Henkilöillä on

tehtäessä, ei päätös perustu pelkästään automaattiseen käsittelyyn tietosuoja-asetuksen näkökulmasta. Euroopan tietosuojatyöryhmä on kuitenkin todennut, ettei rekisterinpitäjä voi kiertää tietosuoja-asetuksen reunaehtoja tekaistulla ihmisen osallistumisella. Ks. 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 22.

⁸ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 5.

⁹ Euroopan parlamentin tutkimus 2019, s. I. Algoritmisen päätöksenteon merkityksen yhteydessä käydyssä keskustelussa sivutaan usein myös algoritmisen päätöksenteon moraalisia ja eettisiä kysymyksiä. Kun ihmistoimijan sijasta päätöksen tekee tekoäly, herää kysymys siitä, millä arvopohjalla tekoäly on ohjelmoitu tekemään päätöksen ja kuka ylipäänsä vastaa päätöksen aiheuttamista seurauksista. Esimerkiksi onnettomuuksien määrä saattaa vähentyä algoritmisen päätöksenteon ansiosta, mutta toisaalta onnettomuuden välttämiseksi tehtävät toimenpiteet nostavat esiin eettisiä kysymyksiä esimerkiksi automaattisesti ohjautuvan ajoneuvon päättäessä suojeleeko se kuljettajan vai ajoväylälle tulleen jalankulkijan terveyttä. Ks. Euroopan parlamentin tutkimus 2019, s. 17–18.

¹⁰ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 5.

¹¹ Euroopan parlamentin tutkimus 2019, s. 20.

¹² Tässä tutkielmassa käytetään jatkossa päätöksenteon kohteena olevan henkilön sijasta termiä rekisteröity. Tietosuoja-asetuksen 4(1)(1) artiklan mukaan rekisteröidyllä viitataan tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön. Rekisteröity-termin käyttö on perusteltua tutkielmassa, koska tutkielmassa käsitellään tilanteita, joissa automaattiseen päätöksentekoon liittyy päätöksenteon kohteena olevan henkilön henkilötietojen käsittelyä.

lisäksi oikeus tietää heidän asiansa käsittelyn perusteet arvioidakseen automaattisen päätöksenteon asianmukaisuutta.¹³ Tämä kuvastaa hyvin sitä tietynlaista jännitettä, joka vallitsee toisaalta asioiden käsittelyn tehostamisen ja toisaalta hallinnon asiakkaiden oikeusturvan kesken.

EU:n tekoälyasetusehdotuksessa kiinnitetään erityistä huomiota olennaisten yksityisten ja julkisten palvelujen ja etujen saatavuuteen ja käyttöön liittyviin tekoälyjärjestelmiin, joita ihmiset tarvitsevat osallistuakseen täysipainoisesti yhteiskuntaan tai parantaakseen elintasoaan. Nämä järjestelmät saattavat vaikuttaa merkittäväällä tavalla alisteisessa asemassa viranomaisiin nähden olevien henkilöiden toimeentuloon ja loukata heidän perusoikeuksiaan, jos tekoälyjärjestelmiä käytetään sen määrittämiseen, tuleeko viranomaisten evätä tällaiset etuudet ja palvelut tai rajoittaa, peruuttaa tai periä ne takaisin.¹⁴

Hallinnon kontekstissa algoritmit asettavat haasteen, sillä niitä ei tällä hetkellä säännellä yksityiskohtaisesti, mutta niitä hyödynnetään kuitenkin vallankäytön apuvälineenä joissakin tapauksissa, eikä niitä tällöin ole mahdollista valvoa samalla tavalla kuin perinteistä valtiojohtoista päätöksentekoa.¹⁵

Algoritmissa päätöksenteon seurauksena voi aiheutua yksilöiden ja ryhmien kannalta kielteisiä vaikutuksia, sillä algoritminen päätöksenteko voi johtaa syrjintään kuten myöhemmin tässä tutkielmassa kuvataan.¹⁶ Syrjintäkiellon tarkoitus on estää erilainen kohtelu ilman hyväksyttävää perustetta ja sitä kautta myös estää mielivaltaisen tai syrjivän päätöksenteon toteutuminen. Yhdenvertaisuusperiaatteen toteutuminen on mahdollista turvata huolehtimalla siitä, että algoritminen päätöksenteko perustuu objektiivisiin, lainsäädännöstä johdettuihin sääntöihin.¹⁷

Algoritmissa päätöksenteon läpinäkymättömyys haittaa syrjivyyden tunnistamista ja sen syiden selvittämistä, sillä sen johdosta on lähes mahdotonta määrittää, aiheuttiko syrjivyyden päätöksentekojärjestelmässä tapahtunut päättelyprosessi.¹⁸ Tämä on seurausta

¹³ Eduskunnan oikeusasiamies, Verohallinnon automatisoitu päätöksentekomenettely ei täytä perustuslain vaatimuksia - Tiedotteet, 2019 <<https://www.oikeusasiamies.fi/fi/-/verohallinnon-automatisoitu-paatoksentekomenettely-ei-tayta-perustuslain-vaatimuksia>> (haettu 11.3.2022).

¹⁴Ehdotus Euroopan parlamentin ja neuvoston asetusta tekoälyä koskevista yhdenmukaistetuista säännöistä (tekoälysäädös) ja tiettyjen unionin säädösten muuttamisesta, COM/2021/206 final, johdanto-osan 37 kappale.

¹⁵ Diakopoulos 2016, s. 58.

¹⁶ Euroopan komission valkoinen kirja tekoälystä 2020, s. 1. On myös huomattava, että algoritminen päätöksenteko voi aiheuttaa laajemmin riskejä hallinnon kontekstissa hallintolain (434/2003) periaatteiden asianmukaisen toteutumisen osalta (esimerkiksi palveluperiaate, neuvontaperiaate ja hyvän kielenkäytön vaatimus). Ks. tarkemmin yllä kuvatuista riskeistä oikeusministeriön arviomuistio (oikeusministeriön arviomuistio 2020, s. 33–34). Arviomuistiossa on kuvattu myös yleisemmin hallinnon automaattisen päätöksenteon riskejä ja sääntelytarpeita.

¹⁷ Oikeusministeriön arviomuistio 2020, s. 32.

¹⁸ O'Hara 2020, s. 1.

algoritmien fundamentaalisesta erosta ihmisjohtoiseen päätöksentekoprosessiin: koneoppimisalgoritmien toiminta perustuu enemmän asioiden väliseen korrelaatioon, kuin ihmisen ajattelulle ominaiseen syy-seuraussuhteeseen. Algoritmit eivät myöskään tuota ihmismielelle ominaisia selityksiä, vaan niiden toiminta perustuu erinäisiin ennusteisiin ja todennäköisyyksien arviointiin.¹⁹ Suurin riski päätöksen syrjivyydelle onkin käsillä tilanteissa, joissa päätöksenteon perusteet eivät ole julkisia.²⁰

Olennaista on tunnistaa se, kenelle päätöksenteon perusteet kohdistetaan (media, valvontaviranomaiset vai yksittäiset rekisteröidyt).²¹ Painopiste tässä tutkielmassa on nimenomaan yksittäistä henkilöä koskevan päätöksenteon syrjivyyden tunnistamisessa päätöksenteon ymmärrettävyyden kautta. Silti arviointiin linkittyy olennaisena osana myös päätöksentekojärjestelmän tarkastelu laajemmin yksittäisen päätöksen lisäksi.²²

Tärkeä kokonaisuus on myös sen tunnistaminen, mitä päätöksenteon ymmärrettävyydellä pyritään saavuttamaan yksittäisen henkilön kannalta. Toisin sanoen kyseeseen tulee sen arviointi, onko kyseessä yksittäisen päätöksen sisältämän syrjinnän jälkikäteinen ”korjaaminen” vai onko tavoitteena tämän ohella vaikuttaminen myös laajemmin päätöksentekojärjestelmän syrjivyyden poistamiseen?²³

Euroopan unionin tasolla tekoälyjärjestelmien läpinäkyvyys on nostettu tärkeäksi asiaksi tekoälyjärjestelmien riskienhallinnan näkökulmasta.²⁴ Lisäksi kansallisella tasolla on valmisteilla hallinnon automaattista päätöksentekoa koskevaa sääntelyä.²⁵ Edellä mainitut tekoälyä koskevat regulatoriset toimenpiteet tulee käsittää osaksi laajempaa viime aikoina käytyä oikeudellista, yhteiskunnallista ja tieteidenvälistä keskustelua tekoälyjärjestelmien etiikasta, johon päätöksentekojärjestelmien avoimuuden ja läpinäkyvyyden vaatimusten voidaan katsoa kuuluvan osaksi.²⁶

¹⁹ Selbst –Barocas 2018, s. 1089–1090.

²⁰ Oikeusministeriön arviomuistio 2020, s. 32.

²¹ Hakkarainen ym. 2020, s. 44.

²² Algoritmisen päätöksenteon laaja-alaisempi ymmärrettävyys yhteiskunnassa ja algoritmisen päätöksenteon valvonta viranomaisten tai muiden valvontaorganien toimesta ovat kiinnostavia tutkimuskysymyksiä, mutta tässä tutkielmassa keskitytään tarkastelemaan algoritmisen päätöksenteon ymmärrettävyyttä nimenomaan yksittäisen rekisteröidyn näkökulmasta.

²³ Hakkarainen ym. 2020, s. 18.

²⁴ Euroopan komission valkoinen kirja tekoälystä 2020, s. 15.

²⁵ Oikeusministeriö, Automaattista päätöksentekoa koskevan hallinnon yleislainsäädännön valmistelu, OM021:00/2020

<<https://oikeusministerio.fi/hanke?tunnus=OM021:00/2020>> Oikeusministeriö>FI>Hankkeet ja säädösvalmistelu>Hanke (haettu 11.3.2022).

²⁶ Hakkarainen ym. 2020, s. 4.

Tutkielman kannalta merkittävin normisto on tietosuojalainsäädäntö, tarkemmin sanottuna EU:n yleinen tietosuojasetus (jäljempänä ”tietosuojasetus”).²⁷ Tutkielmassa pyritään analysoimaan tietosuojasetuksen asettamia velvoitteita suhteessa algoritmisen päätöksenteon asettamiin haasteisiin. Tietosuojasetuksessa säädelään myös aiemmin esille nousutta algoritmisen päätöksenteon läpinäkyvyyttä: tietosuojasetuksessa asetetaan henkilötietoja käsittelevälle algoritmista päätöksentekojärjestelmää hyödyntävälle rekisterinpitäjälle erinäisiä velvoitteita varmistaa, että rekisteröity ymmärtää riittävällä tavalla henkilötietojensa käsittelyn sisältäen algoritmisen järjestelmän toiminnan yksityiskohdat.²⁸

Tietosuojasetuksen velvoitteet ovat kuitenkin osittain tulkinnanvaraisia siltä osin, miten tarkasti rekisterinpitäjän tulee ylipäänsä kertoa algoritmisen järjestelmän toiminnasta.²⁹ Tämä on omiaan heikentämään rekisteröityjen itsemääräämisoikeutta, mutta myös rekisterinpitäjien oikeusvarmuutta.³⁰ Tietosuojalainsäädännön lisäksi hallintolaissa säädetään päätösten perusteluvelvollisuudesta.³¹

Burrell on hahmottanut algoritmisen päätöksenteon läpinäkymättömyyden³² kolmeen eri muotoon, joita ovat (1) läpinäkymättömyys, joka saa alkunsa organisaatioiden haluttomuudesta jakaa algoritmeja koskevaa tietoa, (2) läpinäkymättömyys, joka saa alkunsa siitä, etteivät ihmiset kykene ymmärtämään heille toimitettuja algoritmeja koskevia tietoja ja (3) läpinäkymättömyys, joka saa alkunsa koneoppimisjärjestelmien monimutkaisten toimintaperiaatteiden ja ihmismielen riittämättömän päättelykyvyn välisestä ristiriidasta.³³

Tässä tutkielmassa sovelletaan Burrellin läpinäkymättömyyden jaottelua pyrkimällä ensin määrittelemään ne tiedot, joita rekisterinpitäjien tulee toimittaa rekisteröidyille

²⁷ Euroopan parlamentin ja neuvoston asetus (EU) 2016/679, annettu 27 päivänä huhtikuuta 2016, luonnollisten henkilöiden suojelusta henkilötietojen käsittelyssä sekä näiden tietojen vapaasta liikkuvuudesta ja direktiivin 95/46/EY kumoamisesta. Tietosuojasetuksen suhdetta kansalliseen hallintomenettelyyn on tarkastellut esimerkiksi oikeuskansleri, joka on todennut ratkaisussaan, että ”*Vaikka yleinen tietosuojasetus ei säätele hallintomenettelyä, asettaa se silloin, kun päätöksenteko perustuu henkilötietojen käsittelyyn ja erityisesti pelkääntään automaattiseen henkilötietojen käsittelyyn, vaatimuksia päätöksenteolle.*” Ks. Valtioneuvoston oikeuskansleri, oikeuskanslerin ratkaisu 2021, OKV/131/70/2020 20.4.2021.

²⁸ Tietosuojasetuksen 12–15 artiklat ja 22 artikla.

²⁹ Brkan 2018, sivu 1.

³⁰ Ks. läpinäkyvyyden osa-alueista ja merkityksestä 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 4–6, jossa myös kuvataan, että kyseisten suuntaviivojen tarkoituksena on auttaa rekisterinpitäjää hahmottamaan läpinäkyvyysvelvoitteiden toteuttaminen.

³¹ Hallintolain 45 §.

³² Läpinäkyvyydelle ei ole määritelty yhtä tiettyä merkitystä. Tässä tutkielmassa läpinäkyvyydellä viitataan aineiston tai tiedon toimittamiseen tai saataville saattamiseen rekisteröidylle, kun taas ymmärrettävyydellä viitataan siihen, että rekisteröity voi todella ymmärtää hänelle toimitetun tiedon. Läpinäkyvyyden nojalla toimitettujen tietojen ymmärrettävyys taas saattaa edellyttää tiedon sisällön ja merkityksen selittämistä auki tarkemmin rekisteröidylle. Ks. O’Hara 2020, s. 3. Kuitenkin eräissä kohdin tutkielmassa viitataan läpinäkyvyyteen myös lähdeaineiston käsittämässä merkityksessä, sikäli kuin tämä on tarpeen alkuperäisen lähteen merkityksen kannalta.

³³ Burrell 2016, s. 1–2.

algoritmiseen päätöksentekoon liittyen (kohta 1). Tämän jälkeen tutkielmassa arvioidaan sitä, riittääkö tiedon toimittaminen sellaisenaan takaamaan riittävän ymmärryksen algoritmisesta päätöksenteosta syrjivyyden tunnistamiseksi (kohta 2). Lopuksi tarkastellaan sitä, voidaanko algoritmisia päätöksentekojärjestelmiä todella selittää ymmärrettävällä tavalla rekisteröidylle syrjivyyden tunnistamiseksi (kohta 3).

Edellä mainittujen pohjalta tässä tutkielmassa on tunnistettu seuraavat tutkimuskysymykset:

1. Millä tavoin algoritmisessa päätöksenteossa hyödynnettävissä koneoppimiseen perustuvissa tekoälyjärjestelmissä voi ilmetä syrjivyyttä?
2. Millaisia algoritmisen päätöksenteon yksilötason ymmärrettävyyttä edistäviä velvoitteita tietosuojalainsäädännössä ja hallintolaissa asetetaan rekisterinpitäjälle?
3. Ovatko kyseiset ymmärrettävyyttä koskevat velvoitteet riittäviä takaamaan rekisteröidyn ymmärryksen, jotta hänen on mahdollista tunnistaa algoritmisen päätöksenteon syrjivyyttä?

Tutkimuskysymykset voidaan tiivistää kysymykseksi siitä, ovatko algoritmisen päätöksenteon syrjivyyden tunnistamista koskevat haasteet todella mahdollista ratkaista informoimalla niiden kohteena olevaa henkilöä järjestelmän toiminnasta ja päätöksen perusteluista, vai johtavatko järjestelmien toimintaperiaatteet siihen, että tällaisten tietojen toimittaminen ei ole henkilöiden oikeusturvan kannalta merkityksellistä ja voi jopa pahimmillaan johtaa henkilöitä harhaan.³⁴

Tutkimusaiheen merkittävyyttä ja tarvetta tutkimukselle demonstroivat teknologisen kehityksen mukanaan tuoma tarve algoritmisen päätöksenteon normiston päivittämiselle vastaamaan teknologisen kehityksen synnyttämiä uudenlaisia haasteita yksilöiden oikeuksille ja vapauksille. On kuitenkin tärkeää huomata, että algoritmisen päätöksenteon yhteydessä tunnistetut riskit syrjinnälle ovat olemassa myös ihmisvetoisessa päätöksenteossa, sillä myös ihmismieli muodostaa tietynlaisen mustan laatikon.³⁵ Tämä ei tietenkään sellaisenaan merkitse sitä, etteikö tutkimus algoritmisen päätöksenteon mustan laatikon aiheuttamien syrjinnän riskien osalta ole tarpeen.

Kuten Hakkarainen ym. ovat todenneet, vanhojen periaatteiden nouseminen keskiöön hallinnon automaation tutkimuksessa ei ole yllätys, sillä digitalisaation oikeudellisessa tutkimuksessa törmätään usein ensin tutkimuksellisenä ja käytännöllisenä haasteensa sen

³⁴ Edwards - Veale 2017, s. 65–66.

³⁵ Koivisto 2020, s. 4.

hahmottamiseen, millä tavoin uudet teknologiat haastavat tai mahdollisesti jopa vievät pohjan olemassa olevilta periaatteilta.³⁶ Samalla tavalla tässä tutkielmassa tarkastellaan sitä, miten tehokkaasti läpinäkyvyys laajempaan periaatteena sisältäen hallinnon avoimuuden, onnistuu vastaamaan algoritmisen päätöksenteon mukanaan tuomiin haasteisiin.

Syy sille, miksi tutkielmassa lähestytään tekoälyjärjestelmien syrjivyyden ja läpinäkyvyyden sekä ymmärrettävyyden välistä suhdetta nimenomaan algoritmisen päätöksenteon kannalta johtuu siitä, että tässä tutkielmassa kuvatussa akateemisessa keskustelussa ja lainsäädäntöhankkeissa on pitkälti analysoitu asiaa algoritmisen päätöksenteon (mukaan lukien hallintopäätösten) kautta, eikä niinkään laajemman hallinnon automaation kautta.

Lisäksi on huomattava, että algoritmiseen päätöksentekoon sisältyy rekisteröityä koskeva päätös, kun taas yleisluonteiseen hallinnon prosessien automaatioon ei sisälly vastaavalla tavalla henkilöön vaikuttavaa päätöstä. Hallinnon prosessien automatisoinnin aiheuttamia potentiaalisia syrjinnän riskejä ei kuitenkaan tästä huolimatta tule väheksyä.³⁷ Hallinnon kontekstissa on myös huomattava oikeussuojakeinojen eroavaisuudet automaation ja hallintopäätöksen kesken. Hallintopäätöksestä voi valittaa, mutta muussa hallintotoiminnassa arvioitavaksi tulevat mahdolliset muut oikeussuojakeinot, joita ei kuitenkaan käsitellä tarkemmin tässä tutkielmassa pois lukien tietosuojasetuksen mukaiset rekisteröidyn oikeudet.³⁸

1.2. Tutkimusmetodi ja lähteet

Tutkielma sisältää useita eri oikeustieteellisessä tutkimuksessa sovellettavia metodologioita. Käytän tutkielmassani Hirvosen teoksessa ”*Mitkä menetit? Opas oikeustieteen metodologiaan*” kuvattuja metodologioita.³⁹ Tutkielmassani systematisoin voimassa olevaa oikeutta etenkin tietosuojaoikeuden ja hallinto-oikeuden osa-alueilla, jolloin kyseeseen tulee näiltä osin lainopillinen menetelmä.⁴⁰ Hirvonen on määrittänyt lainopin oikeusnormien ja

³⁶ Hakkarainen ym. 2020, s. 14.

³⁷ Esimerkkinä hallinnon prosessien automaation riskeistä voidaan ajatella tilannetta, jossa henkilö luokitellaan tekoälyjärjestelmän toimesta tiettyyn palvelutarvesegmenttiin, joka johtaa siihen, että henkilöön kohdistetaan tämän tarpeisiin nähden riittämätöntä palvelua, mikä vaikuttaa kielteisesti kyseisen henkilön asian hoitoon (esimerkiksi henkilö ei saa tarvittavaa neuvontaa tai hänen asiansa käsittely viivästyy kohtuuttomasti).

³⁸ Esimerkiksi sosiaalihuollon asiakkaan asemasta ja oikeuksista annetun lain (812/2000) 5 luvun 23 §:ssä säädetään sosiaalihuollon toteuttamiseen liittyvästä muistutuksen tekemisen mahdollisuudesta. Hallinnon automaattisen päätöksenteon säädösvalmistelun yhteydessä on todettu seuraavaa tosiasiallisesta hallintotoiminnasta ”*Esimerkiksi tosiasiallisen hallintotoiminnan automatisointiin liittyy merkittäviä perusoikeuskysymyksiä, joita on syytä selvittää tarkemmin erikseen.*” Ks. oikeusministeriön arviomuistio 2020, s. 62.

³⁹ Hirvonen 2011.

⁴⁰ Hirvonen 2011, s. 45.

niiden ajatussisältöjen tutkimiseksi, jolloin lainopin keskeisen menetelmän muodostaa tulkinta.⁴¹

Tarkemmin sanottuna, tutkielmassani systematisoin ja jäsennän eri oikeudenalojen kuten hallinto-oikeuden ja tietosuojaoikeuden käsitteitä, niiden oikeusperiaatteita sekä niiden teoreettisia rakennelmia. Tutkielmani tulee myös sisältämään oikeusperiaatteiden välistä punnintaa ja tasapainottamista. Tämä ohjaa tutkielmani metodia käytännöllisen lainopin suuntaan.⁴²

Tutkielmani toinen tutkimuskysymys sisältää myös niin sanotun analyttisen oikeustieteen menetelmän mukaista tutkimusta tarkastellessaan objektiivisesti oikeussääntöjä (tietosuoja-asetus) ja kuvatessaan niitä samalla pyrkien havaitsemaan oikeudelliseen ajatteluun sisältyviä epäselvyyksiä (erityisesti tietosuoja-asetuksen läpinäkyvyyden toteuttamiseen liittyviä epäselvyyksiä). Epäselvyyksien havaitsemisen jälkeen tutkielmassa pyritään analysoimaan ja sitä kautta täsmentämään käsitteitä. Tavoitteena on tätä kautta muodostaa tarkempi, selkeämpi ja monipuolisempi kuva oikeudellisesti merkittävästä asiasta, jotta on mahdollista ratkaista oikeusongelma, eli läpinäkyvyyden velvoitteiden epätäsmällisyys algoritmisen päätöksenteon osalta.⁴³

Tutkielmassa tuotettava tieto on kohdistettu pääasiassa muille kuin oikeussalissa toimiville tahoille. Tällöin Minkkisen mukaan tutkimusta ei tulisi mieltää suuntauksellisessa mielessä lainopilliseksi. Minkkisen mukaan tutkimus, joka pyrkii kartoittamaan voimassa olevan lainsäädännön sisältämiä puutteita ja niiden muutostarpeita, sisältää arvoperusteisen kannanoton voimassa olevasta oikeudesta. Tällöin kyseeseen tulisivat hänen mukaansa joskus *de lege ferenda* -väitteiksi kutsutut argumentit, joiden on katsottu lähinnä muodostavan todellista oikeudellista tutkimusta *de lege lata* muodossa täydentäviä oikeuspoliittisia ”kannanottoja”.⁴⁴

Tutkielman olennainen teema muodostuu oikeudellisten normien riittävyuden arvioinnista suhteessa luonnollisten henkilöiden ominaisuuksien (ymmärryskyvyn) asettamiin rajoitteisiin. Vaikka tutkimusongelman ratkaisuksi saattaisi soveltua empiiristen tutkimusmenetelmien tai sosiologiasta peräisin olevien yhteiskuntateoreettisten viitekehyksien käyttö, ei tässä tutkielmassa ole kuitenkaan valittu empiiristä tutkimusmenetelmää. Syy tälle on se, että tutkielmassa hyödynnetty lähdeaineisto sisälsi

⁴¹ Hirvonen 2011, s. 36.

⁴² Hirvonen 2011, s. 25.

⁴³ Hirvonen 2011, s. 46.

⁴⁴ Minkinen 2017, s. 919.

tutkimuksen tarpeiden kannalta riittävällä tavalla jo kerättyä tutkimusaineistoa, jota tutkielmassa oli mahdollisuus analysoida.⁴⁵ Lisäksi aihetta koskevan empiirisen tiedon kerääminen olisi todennäköisesti osoittautunut haastavaksi huomioiden erityisesti tutkittavien henkilöiden tunnistamisen vaikeus sekä tutkielman kannalta riittävän kattavien ja edustavien algoritmisten järjestelmien valikoiminen.

Päätösaunomaation tutkimuksen osalta keskiöön ovat vasta 2000-luvun puolella nousseet klassisiin hallinto-oikeudellisiin kysymyksiin lukeutuvat hyvän tai avoimen hallinnon toteutumisen kysymykset.⁴⁶ Hallinto-oikeuden lisäksi algoritminen päätöksenteko asettuu informaatio-oikeuden alle, jossa esille nousevia teemoja ovat muun muassa oikeudenmukaisuus, tietoturva ja läpinäkyvyys. Hakkarainen ym. ovat todenneet, että informaatio-oikeus ei kuitenkaan perusta riittävää pohjaa hallintoautomaation tutkimiselle, vaan huomioon on myös otettava hallinto-oikeudelliset kysymykset.⁴⁷ Tässä tutkielmassa omaksutaan sekä hallinto-oikeudellinen, että informaatio-oikeudellinen lähestymistapa tutkimuskysymyksiin.

Lisäksi tutkielman teema etenkin ensimmäisen ja kolmannen tutkimuskysymyksen osalta edellyttää myös tietojenkäsittelytieteen termistön ja tutkimuksen hyödyntämistä osana analyysiä. Kyseeseen ei tästä huolimatta tule poikkitieteellinen tutkimus, sillä tutkielmassa ei varsinaisesti hyödynnetä tietojenkäsittelytieteen tutkimusmenetelmiä. Tutkielmaa on mahdollista lähestyä kuvaamalla sen sisältävän laajassa merkityksessä elementtejä oikeus- ja yhteiskuntatieteellisestä tutkimuksesta, sillä tutkielmassa analysoidaan oikeutta sen yhteiskunnallisissa konteksteissa.⁴⁸ Lähdemateriaalin osalta on kuitenkin todettava, että pääosa kirjallisuudesta muodostuu oikeustieteellisestä kirjallisuudesta tietojenkäsittelytieteen muodostaessa täydentävän ja tutkielman teknisempien osien tarpeita palvelevan lähdeaineiston.

Tutkielmassa käsitellään erityisesti tietosuojaa-asetusta, jossa vahvistetaan säännöt luonnollisten henkilöiden suojelulle henkilötietojen käsittelyn yhteydessä sekä henkilötietojen vapaata liikkuvuutta koskevat säännöt. Tietosuojaa-asetus on EU-tason säädös, joka on siten kaikilta osiltaan velvoittava ja jota tulee soveltaa sellaisenaan kaikissa

⁴⁵ Minkkinen 2017, s. 919.

⁴⁶ Hakkarainen ym. 2020, s. 13.

⁴⁷ Hakkarainen ym. 2020, s. 15.

⁴⁸ Minkkinen 2017, s. 922.

EU:n jäsenvaltioissa. EU-lainsäädännön osana tietosuoja-asetus on ensisijainen suhteessa kansallisiin säännöksiin.⁴⁹

Tutkielmassa sivutaan myös kansallista automaattista päätöksentekoa koskevan hallinnon yleislainsäädännön valmistelua.⁵⁰ Lisäksi tutkielmassa käsitellään hallintolakia ja erityisesti sen 45 §:n mukaista päätösten perusteluvelvollisuutta.

Tutkielmassa on käytetty laajasti myös muita kuin niin sanottuja velvoittavia oikeuslähteitä. Olennaisimpia näistä ovat 29 artiklan mukaisen tietosuojatyöryhmän (jäljempänä ”WP29 tietosuojatyöryhmä”) suuntaviivat.⁵¹ Myös kansallisena valvontaviranomaisena toimivan yhdenvertaisuusvaltuutetun ja laillisuusvalvojan toimivan eduskunnan apulaisoikeusasiamiehen päätöksiä hyödynnetään tutkielmassa.⁵²

Hallinnon automatisoinnin on havaittu tuottavan monimuotoisia ja monimutkaisia ilmiöitä. Samalla on kuitenkin havaittavissa, että ongelmien ratkaisu edellyttää oikeustieteen lisäksi muiden tieteenalojen kuten tietojenkäsittelytieteen ja datatieteen panosta.⁵³ Myös tämä tutkielma on hyödyntänyt tämän yhteistyön tuloksena syntyntä tietoa.

Olennaisen osan lähdeaineistoa muodostavat myös oikeuskirjallisuus sekä jossain määrin tietojenkäsittelytieteitä koskeva kirjallisuus. Edellä mainittujen lähteiden käyttöä perusteleen nähdäkseni se, että tutkielman aihepiiri käsittelee verraten tuoretta ilmiötä, josta ei ole vielä

⁴⁹ Ks. Valiokunnan lausunto PeVL 14/2018 vp. HE 9/2018 vp Hallituksen esitys eduskunnalle EU:n yleistä tietosuoja-asetusta täydentäväksi lainsäädännöksi. Tietosuoja-asetuksen ensisijaisuus on olennaista tutkielmassa kuvattujen kansallisten normien yhteydessä tapahtuvan vertailun kannalta, jolloin tulee tiedostaa, että tietosuoja-asetuksen mukaisilla normeilla on mahdollisissa kollisiotilanteissa ensisijaisuus. Tässä tutkielmassa ei kuitenkaan tarkastella normikollisioita tietosuoja-asetuksen ja kansallisten normien välillä.

⁵⁰ Kyseisen oikeusministeriön alaisen hankkeen tarkoituksena on luoda hallinnon yleislainsäädäntöön säännökset, joiden kautta voidaan varmistua hallinnon lainalaisuuden, hyvän hallinnon periaatteiden, oikeusturvan, julkisuusperiaatteen ja virkavastuun toteutumisesta automaattisen päätöksenteon yhteydessä hallinnon kontekstissa. Ehdotuksessa hallintolakiin lisättäisiin hallintoasian automaattisen ratkaisemisen salliva muotoilu, minkä lisäksi julkisen hallinnon tiedonhallinnasta annettuun lakiin (906/2019) (tiedonhallintalaki) lisättäisiin erinäistä sisältöä muun muassa automatisoidun toimintaprosessin laadunvarmistamisesta ja käyttöönnotosta. Lain on tarkoitus tavoitteen mukaan tulla voimaan 1.1.2023. Ks. Oikeusministeriö, Automaattista päätöksentekoa koskevan hallinnon yleislainsäädännön valmistelu, OM021:00/2020 <<https://oikeusministerio.fi/hanke?tunnus=OM021:00/2020>> Oikeusministeriö>FI>Hankkeet ja säädösvalmistelu>Hanke (haettu 11.3.2022).

⁵¹ WP29 tietosuojatyöryhmä oli riippumaton EU:n työryhmä, jonka tehtävänä oli käsitellä yksilöiden suojelua henkilötietojen käsittelyssä koskevia kysymyksiä 25. toukokuuta 2018 asti. Ks. *Euroopan tietosuojaneuvosto*, Tietosuojatyöryhmä WP 29 <<https://edpb.europa.eu/about-edpb/more-about-edpb/article-29-working-party-fi>>Home>About EDPB>Tietosuojatyöryhmä WP 29 (haettu 15.3.2022). Nykyisin toiminnassa olevan Euroopan tietosuojaneuvoston suuntaviivat ja suositukset eivät ole oikeudellisesti sitovia, mutta ne kuvastavat valvontaviranomaisten näkemyksiä. Ks. Euroopan tietosuojaneuvosto, EDPB Statement : EDPB cooperation on the elaboration of guidelines, 2021 <<https://edpb.europa.eu/news/news/2021/edpb-statement-edpb-cooperation-elaboration-guidelines-en>>Home>News>EDPB Statement: EDPB cooperation on the elaboration of guidelines (haettu 15.3.2022).

⁵² Yhdenvertaisuusvaltuutettu on itsenäinen ja riippumaton viranomainen, jonka tehtävänä on edistää yhdenvertaisuutta ja puuttua syrjintään. Valtuutetulla on myös eräitä raportointi- ja valvontavelvoitteita. Lisäksi valtuutetun tehtäviin kuuluu ulkomaalaisten oikeuksien ja aseman edistäminen. Ks. Yhdenvertaisuusvaltuutettu, yhdenvertaisuusvaltuutettu <<https://syrjinta.fi/etusivu>> (15.3.2022). Oikeusasiamies ja apulaisoikeusasiamiehet vastaavat laillisuusvalvonnasta. Ks. Eduskunnan oikeusasiamies, oikeusasiamies ja apulaisoikeusasiamiehet <<https://www.oikeusasiamies.fi/fi/oikeusasiamies-ja-apulaisoikeusasiamiehet>> (haettu 15.3.2022).

⁵³ Koulou 2018, s. 865.

ehtinyt muodostua oikeuskäytäntöä, eikä tietosuojasetuksessa ole saatavilla aihetta koskien valmistelumateriaalia tukemaan analyysiä.

1.3. Keskeisiä rajauksia

Tutkielmani käsittelee useita eri oikeudenaloja, joista merkittävimmän osan muodostavat tietosuojalainsäädäntö ja hallintolainsäädäntö. Tutkielmassa sivutaan myös syrjintäkieltoa ja yhdenvertaisuutta koskevia lakeja, mutta niiden tarkempi analysointi jää tämän tutkielman ulkopuolelle. Rajaus on nähdäkseni perusteltu tutkielmalle asetettujen pituusvaatimusten, mutta sen puolesta puhuu myös se, että tutkielman tarkoituksena ei ole arvioida syrjintäkiellon alaisen syrjinnän kriteereiden täyttymistä algoritmisen päätöksenteon yhteydessä.

Tutkielmassa ei ole pyritty tästä johtuen kuvaamaan tyhjentävästi kaikkia syrjintäsäännöksiä, vaan siinä on keskitytty tutkielman kannalta olennaisimpiin säännöksiin. Jatkotutkimuksen kannalta voisi kuitenkin olla tarpeen ja hyödyllistä tarkastella eri syrjintäperusteita tarkemmin algoritmisen päätöksenteon syrjivyyden tunnistettavuuden yhteydessä.

Painopiste tutkielmassa on koneoppimiseen perustuvissa järjestelmissä. Sen sijaan tässä tutkielmassa ei keskitytä sääntöpohjaisiin päätöksentekojärjestelmiin, sillä niiden osalta päätöksenteon ymmärrettävyys ei pääpiirteissään näyttäyty yhtä monimutkaisena, kuin koneoppimiseen perustuvien järjestelmien osalta.⁵⁴

Tässä tutkielmassa tarkastellaankin algoritmista päätöksentekoa nimenomaan julkisen sektorin näkökulmasta, mutta havainnot ovat eräitä hallinnon normeihin liittyviä löydöksiä lukuun ottamatta käyttökelpoisia myös yksityisen sektorin kontekstissa.

Kansallisessa lainsäädännössä on määritelty useita eri henkilöön liittyviä syitä, joiden nojalla ketään ei saa syrjiä. Näitä ovat muun muassa henkilön ikä, alkuperä, kieli, uskonto, perhesuhteet, terveydentila ja seksuaalinen suuntautuminen.⁵⁵ Lisäksi sukupuoleen perustuva syrjintä on kielletty.⁵⁶ Tässä tutkielmassa on päädytty käsittelemään yleisesti syrjintää kiinnittymättä tarkemmin mihinkään tiettyyn syrjintäperusteeseen. Tutkielmassa kuitenkin korostuvat jossakin määrin sukupuolta, alkuperää ja ikää koskevat mahdolliset

⁵⁴ Goldenfein 2019, s. 42.

⁵⁵ Yhdenvertaisuuslain 8 §:ssä luetellaan eri henkilöön liittyviä syitä, joiden nojalla tapahtuva syrjintä on kiellettyä.

⁵⁶ Sukupuoleen perustuvan syrjinnän kiellosta ja sukupuolten välisen tasa-arvon edistämisestä säädetään tasa-arvolaissa.

syrjintäperusteet, sillä erityisesti näitä henkilön ominaisuuksia on käsitelty lähdemateriaalissa.

Tarkasteltaessa hallinnon hyödyntämien algoritmisten päätöksentekojärjestelmien mahdollisen syrjivyyden tunnistettavuutta, on tarpeen arvioida tekoälyjärjestelmien teknisiä toiminnallisuuksia. Tässä tutkielmassa ei ole kuitenkaan tarkoitus kuvata tyhjentävästi tekoälyjärjestelmien teknisiä yksityiskohtia, mutta yleinen kuvaus algoritmisen päätöksenteon teknisestä toteutustavasta luo pohjan tutkielmassa esitettävälle oikeudelliselle analyysille.

Silloin kun on kyse hallinnon organisaation tekemästä hallintopäätöksestä, joka täyttää tietosuoja-asetuksessa automaattiselle päätöksenteolle asetetut kriteerit, tulevat päätöksentekoon sovellettavaksi hallinto-oikeudelliset ja tietosuoja-asetuksen mukaiset velvoitteet.⁵⁷ Tosiasiallisen hallintotoiminnan ja hallinnollisen sääntelyn tai muun yksipuolisen julkisen vallan käytön välinen ero on suhteellinen.⁵⁸ Lisäksi on huomattava, että tietosuoja-asetuksessa ei erotella hallintopäätöksiä ja tosiasiallista hallintotoimintaa automaattisen päätöksenteon osalta. Tässä tutkielmassa ei kuitenkaan arvioida tarkemmin rajanvetoa hallintopäätösten ja tosiasiallisen hallintotoiminnan välillä. Tässä tutkielmassa ei myöskään ole tarkoitus ottaa kantaa siihen, pitäisikö hallinnon organisaatioiden mahdollisuuksia ottaa käyttöön algoritmisia päätöksentekojärjestelmiä rajata.

Rekisteröidyn oikeuksien toteuttaminen algoritmisen päätöksenteon yhteydessä asettaa haastavia ja mielenkiintoisia kysymyksiä esimerkiksi henkilötietojen poistamisen tai henkilötietojen käsittelyn rajoittamisen toteutusmahdollisuuksien kannalta. Näitä teemoja ei kuitenkaan käsitellä tässä tutkielmassa tarkemmin.⁵⁹ Lisäksi viranomaisten asiakirjojen julkisuutta koskeva sääntely on noussut esille akateemisessa keskustelussa järjestelmien lähdekoodin yhteydessä, mutta tässä tutkielmassa ei tarkastella tarkemmin algoritmisten päätöksentekojärjestelmien dokumentaatiota viranomaisten asiakirjojen julkisuutta koskevan sääntelyn kannalta.⁶⁰

1.4. Tutkielman rakenne

⁵⁷ Tietosuoja-asetuksen 22 artiklassa säädetään automaattisesta päätöksenteosta ja hallintolaissa hallintopäätöksistä ja muusta hallinnosta. Ks. tarkemmin tähän liittyen luku 2.

⁵⁸ Mäenpää, hallinto-oikeus, sähköinen teos, II yleinen hallinto-oikeus, 4. julkishallinnon tehtävät ja toimintamuodot, viranomaisen toimintamuodot, tosiasiallinen hallintotoiminta, palvelujen toteuttamista ei-kirjallisessa muodossa.

⁵⁹ Ks. tähän liittyen Janssen 2020, s. 87.

⁶⁰ Ks. tähän liittyen Hakkarainen ym. 2020.

Tutkielmani koostuu neljästä pääasiallisesta luvusta. Ensimmäisessä pääasiallisessa luvussa (algoritmisen päätöksenteon toteutus ja syrjivyyden ilmeneminen järjestelmissä) käsittelen yleisellä tasolla algoritmisen päätöksenteon toteutusta tekniseltä kannalta tarkasteltuna. Lisäksi luvussa kuvataan hallinnon päätöksentekoa pääpiirteissään. Lopuksi luvussa käsitellään myös syrjivyyden syitä ja ilmenemismuotoja algoritmisen päätöksenteon yhteydessä.

Toisessa pääasiallisessa luvussa (taustalla vaikuttava lainsäädäntö) pyrin kuvaamaan algoritmisen päätöksenteon syrjivyyteen ja syrjintään liittyvät oikeuslähteet tutkielman kannalta olennaisilta osin. Luku toimii taustoittavana osiona seuraaville luvuille.

Kolmannessa pääasiallisessa luvussa (algoritmisen päätöksenteon läpinäkyvyys ja ymmärrettävyys) kuvaan läpinäkyvyyden käsitettä sekä laajasti ymmärrettynä, että tietosuojasetuksen läpinäkyvyyden periaatteen kautta tarkastellen. Lisäksi käsittelen tietosuojasetuksesta mahdollisesti johdettavissa olevaa selitysoikeutta sekä päätösten perusteluvollisuutta. Lopuksi analysoin algoritmista päätöksentekoa koskevien tietojen toimittamista rekisteröidylle suhteessa päätöksenteon elinkaaren asettamiin reunaehtoihin.

Neljännessä pääasiallisessa luvussa (haasteet algoritmisen päätöksenteon syrjivyyden tunnistamisessa) vedän yhteen edellisissä luvuissa kuvattuja havaintoja rekisteröidylle toimitettavien tietojen tosiasiallisesta merkittävydestä syrjivyyden tunnistamisen kannalta lähestyen asiaa jälleen päätöksenteon elinkaaren kautta.

2. ALGORITMISEN PÄÄTÖKSENTEON TEKNINEN TOTEUTUS JA SYRJIVYYDEN ILMENEMINEN TEKOÄLYJÄRJESTELMISSÄ

2.1. Algoritmisen päätöksenteon tekninen toteutus

Algoritmisen päätöksenteon lainsäädännöllisesti olennaisimman sääntelykehikon tämän tutkielman kannalta muodostaa tietosuoja-asetuksen automaattista päätöksentekoa ja käsittelyn läpinäkyvyyttä koskeva sääntely.⁶¹ Tämän lisäksi hallinnon kontekstissa on tarpeen erotella toisistaan ensinnäkin tosiasiallinen hallintotoiminta ja toisaalta hallintopäätökset alla kuvatulla tavalla.

Tekoälyä on kuvattu koneiden kyvykkyydeksi suorittaa tehtäviä, jotka edellyttävät ihmisen älykkyyttä. Tarkemmin sanottuna tekoäly viittaa tietojärjestelmiin, jotka kykenevät suoriutumaan ihmisen älykkyyttä vaativista tehtävistä, kuten visuaalisesta tunnistamisesta, puheentunnistuksesta tai päätöksenteosta.⁶²

Tekoäly voidaan jaotella niin sanottuun heikkoon ja vahvaan tekoälyyn. Heikko tekoäly on koulutettu suorittamaan tiettyjä rajattuja tehtäviä kuten kasvojen tunnistamista, kun taas yleinen tai vahva tekoäly pyrkii ajan kanssa ohittamaan ihmisen useilla eri toiminnan aloilla.⁶³

Tekoälyjärjestelmien toiminnan kannalta olennaisia ovat algoritmit, joita voidaan kuvata prosesseina, joiden tarkoituksena on ratkaista määritelty ongelma ennalta ohjelmoitujen ohjeiden tai sääntöjen pohjalta. Algoritmit toteuttavat ohjeita ja sääntöjä syöttötietojen, eli datan avulla ja palauttavat tulosteen, eli ratkaisun ongelmaan.⁶⁴ Tarkemmin teknisesti kuvattuna algoritmi tarkoittaa vaiheittaista, ohjelmointikielistä tai matemaattista kuvausta tai ohjetta tavasta, miten järjestelmän tulee suorittaa sille annettu tehtävä.⁶⁵

Seuraavaksi kuvataan tarkemmin koneoppimista, sillä se muodostaa tämän tutkielman kannalta olennaisen tekoälyjärjestelmän toteutustavan.

2.1.1. Koneoppiminen

⁶¹ Ks. erityisesti tietosuoja-asetuksen artiklat 12–15 ja 22.

⁶² Ks. Lexico.com, artificial intelligence, saatavilla: <https://www.lexico.com/definition/artificial_intelligence> (haettu 15.3.2022).

⁶³ Euroopan parlamentin tutkimus 2019, s. 4.

⁶⁴ Euroopan parlamentin tutkimus 2019, s. 3. Algoritmien määritelmän merkitys on nostettu esille lainsäädäntövaiheessa. Ks. valiokunnan lausunto PeVL 7/2019 vp - HE 18/2019 vp Hallituksen esitys eduskunnalle laiksi henkilötietojen käsittelystä maahanmuuttohallinnossa ja eräiksi siihen liittyviksi laeiksi, jossa todetaan sivulla 10: ”Valiokunta toistaa nämä huomiot ja painottaa myös, että algoritmin julkisuuden asianmukainen toteutuminen yksityiselle ymmärrettävässä muodossa edellyttää, että laissa on tarkkarajaisesti ja täsmällisesti määritelty, mitä algoritmilla automatisoidussa päätöksenteossa tarkoitetaan.”

⁶⁵ Hakkarainen ym. 2020, s. 9.

Koneoppimiselle on olemassa useita määritelmiä, joista yhtenä käyttökelpoisena voidaan mainita seuraava Euroopan parlamentin raportissa omaksuttu määritelmä ” *Machine learning is an AI component that provides systems with the ability to automatically learn over time, generally from large quantities of data.*” Koneoppimisjärjestelmän on mahdollista havaita sille annetusta suuresta aineistosta säännönmukaisuuksia, joiden pohjalta se voi laatia ennustuksia. Koneoppimisalgoritmi luo sille syötetyn datan perusteella toisen algoritmin (tätä nimitetään tyypillisesti malliksi).⁶⁶

Koneoppiminen hyödyntää siis tilastotieteen menetelmiä, jonka avulla tietokoneella on mahdollisuus oppia asioita datasta ilman koneen ohjelmointia. Koneoppimista hyödynnetään etenkin tapauksiin, joissa on tarpeen saada käyttöön ennusteita ja luokitella ilmiöiden tai eri toimien aiheuttamia lopputulemia.⁶⁷

Koneoppiminen voidaan kategorisoida eri tyypeihin, joita ovat ohjattu oppiminen, ohjaamaton oppiminen ja vahvistusoppiminen.

1. Ohjatussa oppimisessa mallia opetetaan merkityllä datalla (ns. labelled data). Tämän jälkeen mallin tehtävänä on tunnistaa vastaavanlaisen merkinnän sisältävä data sille syötetyn alkuperäisen opetusdatan ulkopuolisesta datasta. Ohjattua oppimista hyödynnetään esimerkiksi jatkuvien arvojen ennustamiseen, regressioon tai luokittelun kautta toteutettavaan kategorisen muuttujan ennustamiseen.
2. Ohjaamattomassa oppimisessä ei hyödynnetä merkittyä dataa, vaan malli tunnistaa automaattisesti kaavamaisuuksia ja rakenteita opetusdatasta.
3. Vahvistusoppimisella viitataan koneoppimiseen, jossa malli hyödyntää ympäristöltään saamaansa palautetta. Algoritmin tavoitteena on löytää toiminta, joka tuottaa suurimman määrän palkkiota.⁶⁸

Koneoppimisen mallilla viitataan kaikkeen siihen, jonka algoritmi on oppinut juoksutettuaan opetusdatan läpi ja kattaa säännöt, numerot ja muut algoritmiin liittyvät datarakenteet ennustusten tekemiseen. Esimerkkejä koneoppimisalgoritmien malleista ovat lineaarisen regression algoritmit, päätöspuualgoritmit (koostuvat ”jos-niin” tyyppisistä toteamuksista) ja neuroverkkoalgoritmit.⁶⁹

⁶⁶ Euroopan parlamentin tutkimus 2019, s. 4.

⁶⁷ Valtioneuvoston loppuraportti 2019, s. 6.

⁶⁸ Euroopan parlamentin tutkimus 2019, s. 4.

⁶⁹ Machinelearningmastery.com, Difference Between Algorithm and Model in Machine Learning <<https://machinelearningmastery.com/difference-between-algorithm-and-model-in-machine-learning/>> (haettu 15.3.2022).

Oppivat algoritmit edellyttävät opetusdatan avulla toteutettavaa koulutusvaihetta, mitä seuraa algoritmin testaaminen testausdatan avulla. Testausvaihe on olennainen algoritmiin syntyneiden virheiden havaitsemiseen (erityisesti virheellinen luokittelu).⁷⁰

Neuroverkot on mahdollista lukea osaksi koneoppimista.⁷¹ Neuroverkko tarkoittaa tietokonearkkitehtuuria, jossa lukuisat prosessorit ovat yhteen liitettynä ihmisen aivoissa olevien neuronien kaltaisella tavalla ja kykenevät oppimaan yrityksen ja erehdyksen kautta.⁷²

2.2. Profiloinnin ja automaattisen päätöksenteon määritelmä EU:n tietosuojalainsäädännössä

2.2.1. Tietosuoja-asetuksen mukainen automaattinen päätöksenteko

Tietosuoja-asetuksen johdanto-osassa todetaan, että ”rekisteröidyillä olisi oltava oikeus olla joutumatta hänen henkilökohtaisia ominaisuuksiaan arvioivan, mahdollisesti toimenpiteen sisältävän päätöksen kohteeksi, joka on tehty yksinomaan automaattisen tietojenkäsittelyn perusteella, josta hänelle aiheutuu oikeudellisia vaikutuksia tai joka vaikuttaa häneen vastaavalla merkittävällä tavalla, ilman ihmisen osallistumista.”⁷³

Tietosuoja-asetuksessa ei ole kuitenkaan määritelty teknisessä mielessä tarkemmin automaattista päätöksentekoa. Tietosuoja-asetuksen johdanto-osassa luetellaan esimerkkeinä automaattisesta päätöksenteosta online-luottihakemuksen automaattinen epääminen tai sähköisen rekrytoinnin käytännöt ilman ihmisen osallistumista.⁷⁴

Euroopan tietosuojatyöryhmä on ohjeistuksessaan todennut, että automaattisella päätöksenteolla viitataan päätöksentekoon, jossa päätökset voidaan tehdä täysin automaattisesti ilman ihmisen osallistumista päätöksentekoprosessiin. Euroopan tietosuojatyöryhmän ohjeistuksessa annetaan esimerkkinä automaattisesta päätöksenteosta tilanne, jossa ylinopeussakko määrätään nopeusvalvontakameroiden keräämän todistusaineiston pohjalta ilman ihmisen osallistumista päätöksentekoprosessiin.⁷⁵

⁷⁰ Hildebrandt – Gutwirth 2008, s. 70.

⁷¹ Valtioneuvoston loppuraportti 2019, s. 6.

⁷² Merriam-Webster, neural network, saatavilla: <<https://www.merriam-webster.com/dictionary/neural%20network>> (haettu 16.3.2022). Syviin neuroverkkoihin perustuvat algoritmit saattavat pitää sisällään miljoona sisäistä parametria, eli kerrointa. Tästä seurauksena on lähestulkoon täysi mahdottomuus muodostaa algoritmia ja sen toimintaa koskevaa selkeää kokonaiskuvaa ihmiselle ymmärrettävällä tavalla. Ks. Valtioneuvoston loppuraportti 2019, s. 37.

⁷³ Tietosuoja-asetuksen johdanto-osan 71 kappale.

⁷⁴ Tietosuoja-asetuksen johdanto-osan 71 kappale.

⁷⁵ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 8.

Tietosuoja-asetuksen automaattista päätöksentekoa koskevat velvoitteet soveltuvat siis vain tietyt tunnusmerkistöt täyttävään päätöksentekoon. Tietosuoja-asetuksessa automaattinen päätöksenteko kattaa päätökset, jotka perustuvat pelkästään automaattiseen käsittelyyn. Tällä tarkoitetaan sitä, että päätöksentekoprosessiin ei osallistu ihmistä. Euroopan tietosuojatyöryhmä on todennut, että rekisterinpitäjä ei voi kiertää tietosuoja-asetuksen reunaehtoja tekaistulla ihmistoimijan osallistumisella prosessiin. Esimerkiksi tilanteissa, joissa ihmistoimija soveltaa rutiininomaisesti automaattisesti luotuja profiileita henkilöihin ilman vaikutuksia päätöksen lopputulokseen, kyseeseen tulee pelkästään automaattiseen käsittelyyn perustuva päätös. Ihmisen osallistuminen edellyttää sitä, että päätöksen valvonnalla on merkitystä eikä kyseeseen tule tyhjä ele. Valvonnan suorittavalla taholla tulee olla valtuudet ja pätevyys päätöksen muuttamiseksi ja sen tulee huomioida kaikki asian kannalta merkitykselliset tiedot.⁷⁶ Tässä tutkielmassa tarkastellaan siten myös tilanteita, joissa ihmistoimija soveltaa rutiininomaisesti automaattisesti tuotettuja päätöksiä vaikuttamatta itse päätöksen lopputulokseen.

Lisäksi päätöksen vaikutusten tulee ylittää tietty kynnyks, jotta tietosuoja-asetuksen automaattista päätöksentekoa koskevat normit tulevat kyseeseen. Tämä tapahtuu, mikäli päätöksellä on oikeusvaikutuksia. Oikeusvaikutus edellyttää, että päätös vaikuttaa henkilön laillisiin oikeuksiin, kuten oikeuteen ryhtyä oikeustoimiin. Oikeusvaikutuksena pidetään myös vaikutusta henkilön oikeudelliseen asemaan tai hänen sopimusperusteisiin oikeuksiinsa. Tähän kategoriaan kuuluvat päätökset, jotka johtavat esimerkiksi sopimuksen peruuttamiseen tai lakisääteisen sosiaalietuuden myöntämiseen tai epäämiseen.⁷⁷

Päätöksen vaikutus ylittää myös vaaditun kynnyksen, mikäli se vaikuttaa rekisteröityyn vastaavalla tavalla merkittävästi, vaikkei päätöksentekoprosessi vaikuttaisi henkilöiden laillisiin oikeuksiin. Vaikutusten on kuitenkin oltava riittävän suuria tai huomattavia, jotta tietojenkäsittelyn vaikutukset henkilöön ovat merkittäviä. Tämä tarkoittaa, että päätöksen on osaltaan kyettävä mahdollisesti vaikuttamaan merkittävästi henkilöiden olosuhteisiin, käyttäytymiseen tai valintoihin, tai luomaan pitkäaikaisia tai pysyviä vaikutuksia henkilöille tai johtaa henkilöiden syrjäytymiseen tai syrjintään.⁷⁸

2.2.2. Profilointi

⁷⁶ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 22.

⁷⁷ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 22–23.

⁷⁸ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 23.

Luonnollisia henkilöitä koskevien päätösten tukena hyödynnetään usein henkilöitä koskevia taustatietoja. Yksi tapa muodostaa tällaista taustatietoa on profilointi, joka tarkoittaa menettelyä, joka sisältää tilastollisia päätelmiä. Kyseessä on henkilöitä koskevien ennusteiden tekeminen hyödyntämällä eri lähteistä peräisin olevia tietoja. Tavoitteena on tehdä päätelmiä tietystä henkilöstä muiden tilastolliselta kannalta samanlaisten henkilöiden ominaisuuksien perusteella.⁷⁹

Profiloinnissa on kyse induktiivisesta tavasta tuottaa tietoa arvioimalla korrelaatiosuhteiden kautta sen todennäköisyyttä, että jokin asia tapahtuu samalla tavalla tulevaisuudessa. Profiloinnissa olennaista ei tyypillisesti ole syy-seuraussuhteiden tai korrelaatioiden syiden tunnistaminen, vaan luotettavien ennusteiden laatiminen. Profiloinnin tuottamia profiileja voidaankin luonnehtia tietynlaisiksi hypoteeseiksi.⁸⁰

Kaikki henkilöiden luokittelu ei kuitenkaan täytä tietosuoja-asetuksen profiloinnin määritelmiä. Esimerkiksi yksinkertainen henkilöiden luokittelu heidän tunnettujen ominaisuuksien kuten iän, sukupuolen ja pituuden perusteella ei välttämättä täytä profiloinnin määritelmää. Luokittelun tarkoituksen perusteella voidaan määrittää, onko kyse profiloinnista.⁸¹

Voidaankin todeta, että laajasti ymmärrettynä profilointi on tietojen keräämistä henkilöstä ja henkilön ominaisuuksien tai käyttäytymismallien arvioimista, jotta henkilö voidaan sijoittaa tiettyyn luokkaan tai ryhmään. Tämä mahdollistaa analysoinnin ja ennustamisen esimerkiksi liittyen henkilön kykyyn suorittaa tehtäviä, henkilön kiinnostuksen kohteisiin tai tämän todennäköiseen käyttäytymiseen.⁸²

On myös huomattava, että päätökset, jotka eivät ole kokonaisuudessaan automaattisia, saattavat pitää sisällään profilointia. Esimerkiksi pankki voi ennen lainan myöntämistä huomioida lainaa hakevan luottopisteistyksen, minkä jälkeen pankin virkailijat kuitenkin suorittavat merkityksellisiä täydentäviä toimia, ennen kuin päätös tehdään.⁸³

Profiloinnin eri ilmenemismuotoja voidaan havainnollistaa seuraavalla jaottelulla:

⁷⁹ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 7. Tietosuoja-asetuksen johdanto-osan 71 kappaleen mukaan profilointi tarkoittaa henkilötietojen automaattista käsittelyä henkilön henkilökohtaisten ominaisuuksien arvioimiseksi siltä osin kuin arvioimisella on rekisteröityyn kohdistuvia oikeudellisia vaikutuksia tai se vaikuttaa rekisteröityyn vastaavasti merkittävällä tavalla. Arvioiminen voi koskea erityisesti rekisteröidyn työsuorituksen, taloudellisen tilanteen, terveyden, henkilökohtaisten mieltymysten tai kiinnostuksen kohteiden, luotettavuuden tai käyttäytymisen, sijainnin tai liikkumisen analysoimista ja ennakoimista.

⁸⁰ Hildebrandt – Gutwirth 2008, s. 18.

⁸¹ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 7.

⁸² 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 8.

⁸³ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 9.

- a) Yleinen profilointi
- b) Päätöksenteko, joka perustuu profilointiin
- c) Ainoastaan automatisoitu päätöksenteko (profilointi mukaan luettuna), joka sisältää rekisteröityä koskevia oikeusvaikutuksia tai joka vaikuttaa häneen vastaavalla tavalla merkittävästi (tietosuoja-asetuksen 22(1) artikla)⁸⁴

Tämän tutkielman kannalta olennainen on c-kohdan mukainen pelkästään automatisoitu päätöksenteko, joka sisältää myös profiloinnin.

2.3. Hallintopäätökset ja tosiasiallinen hallintotoiminta

Hallintopäätös tarkoittaa viranomaisen tekemää ratkaisua hallintoasiassa. Hallintopäätöksessä voidaan asettaa kiello tai rajoitus tai myöntää etu tai oikeus. Hallintopäätöksen tai sen välittömien oikeusvaikutusten kohteena on asianosainen. Hallintopäätöksiin sisältyy oikeudellisesti sitova ratkaisu asianosaisen oikeudesta tai velvollisuudesta.⁸⁵

Hallintopäätösten tuottamat oikeusvaikutukset ovat luonteeltaan vaihtelevia. Hallintopäätökset on mahdollista kategorisoida etuja suoviin (kuten hallintolupiin, etuuksiin ja tukiin), oikeuksia perustaviin (oikeusasemaa koskeva perustaminen tai sen vahvistaminen) ja velvoitteita asettaviin (kuten rajoituksiin, kieltoihin, veroihin ja maksuihin).⁸⁶ Hallintoasian ratkaisua tai sen tutkimatta jättämistä koskevaan päätökseen on mahdollista hakea muutosta. Valittamalla ei kuitenkaan saa hakea muutosta päätökseen, joka koskee vain asian valmistelua tai täytäntöönpanoa.⁸⁷

Tosiasiallisella hallintotoiminnalla viitataan viranomaisen toimintaan, jossa ei tehdä varsinaisesti kirjallisia hallintopäätöksiä. Tällaista toimintaa on esimerkiksi potilaan hoitaminen tai koululaisten opettaminen sekä muiden vastaavien julkisten palvelujen toteuttaminen.⁸⁸

Tosiasiallisen hallintotoiminnan kautta ei suoranaisesti puututa yksityisten oikeuksiin tai velvollisuuksiin, mutta toiminnalla on silti suuri merkitys esimerkiksi sosiaalisten ja

⁸⁴ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 9.

⁸⁵ Mäenpää, hallinto-oikeus, sähköinen teos, II yleinen hallinto-oikeus, 4. julkishallinnon tehtävät ja toimintamuodot, viranomaisen toimintamuodot, hallintopäätökset, hallintopäätöksen peruspäätteet.

⁸⁶ Mäenpää, hallinto-oikeus, sähköinen teos, II yleinen hallinto-oikeus, 4. julkishallinnon tehtävät ja toimintamuodot, viranomaisen toimintamuodot, hallintopäätökset, hallintopäätöksen peruspäätteet.

⁸⁷ Laki oikeudenkäynnistä hallintoasioissa (808/2019), 6 §.

⁸⁸ Mäenpää, hallinto-oikeus, sähköinen teos, II yleinen hallinto-oikeus, 4. julkishallinnon tehtävät ja toimintamuodot, viranomaisen toimintamuodot, tosiasiallinen hallintotoiminta, palvelujen toteuttamista ei-kirjallisessa muodossa.

sivistyksellisten oikeuksien käytännön toteuttamisen edellytysten luomiselle ja elinkeinotoiminnalle.⁸⁹

On tärkeää huomata, että tosiasiallisen hallintotoiminnan ja hallinnollisen sääntelyn tai muun yksipuolisen julkisen vallan käytön välinen ero on suhteellinen. Esimerkiksi hoitotoimenpiteellä on mahdollista rajoittaa yksilön itsemääräämisoikeutta. Tästä syystä myös tosiasiallisessa hallinnossa tulee huomioida oikeusturvan, julkisuuden, yleisen vaikuttamisen ja julkisen valvonnan mahdollisuuksien sekä yleisten hallinto-oikeudellisten periaatteiden toteutuminen.⁹⁰

Silloin kun on kyse hallinnon organisaation tekemästä hallintopäätöksestä, joka täyttää tietosuoja-asetuksessa automaattiselle päätöksenteolle asetetut kriteerit, tulevat päätöksentekoon sovellettavaksi hallinto-oikeudelliset ja tietosuoja-asetuksen mukaiset velvoitteet.⁹¹ On kuitenkin tärkeää huomata, että tietosuoja-asetuksessa ei säädetä hallintopäätöksistä tai tosiasiallisesta hallintotoiminnasta, vaan automaattisen päätöksenteon sääntelyn sovellettavaksi tuleminen määritellään tietosuoja-asetuksessa kuvatulla tavalla. Tässä tutkielmassa ei kuitenkaan arvioida tarkemmin rajanvetoa hallintopäätösten ja tosiasiallisen hallintotoiminnan välillä.

2.3.1. Hallintopäätösten ja tosiasiallisen hallintotoiminnan automatisoinnin nykytila

Hallintopäätösten automatisoinnin osalta nykytilanteessa organisaatioissa hyödynnetään pitkälti sääntöperusteista automaatiota. Suurin osa tällä hetkellä kokonaan automatisoiduista päätöksistä on hyvin massaluonteisia ja yksinkertaisia päätöksiä, joissa asia on mahdollista ratkaista kuulematta asianosaista. Lisäksi päätösten perusteet ovat selkeitä ja laskennallisia, ja ne ovat suoraan laista johdettavissa.⁹²

Myös tosiasiallisen hallintotoiminnan osalta on tunnistettu tarpeita automaation hyödyntämiselle. Koneoppimiseen perustuvien tekoälyjärjestelmien osalta tosiasiallisen

⁸⁹ Mäenpää, hallinto-oikeus, sähköinen teos, II yleinen hallinto-oikeus, 4. julkishallinnon tehtävät ja toimintamuodot, viranomaisen toimintamuodot, tosiasiallinen hallintotoiminta, palvelujen toteuttamista ei-kirjallisessa muodossa.

⁹⁰ Mäenpää, hallinto-oikeus, sähköinen teos, II yleinen hallinto-oikeus, 4. julkishallinnon tehtävät ja toimintamuodot, viranomaisen toimintamuodot, tosiasiallinen hallintotoiminta, palvelujen toteuttamista ei-kirjallisessa muodossa.

⁹¹ Tietosuoja-asetuksen 22 artiklassa säädetään automaattisesta päätöksenteosta ja hallintolaissa hallintopäätöksistä ja muusta hallinnosta.

⁹² Oikeusministeriön arviomuistio 2020, s. 15. Sääntöpohjainen automaattinen päätöksenteko tai päätösaunomaatio tarkoittaa sitä, että päätöksessä sovellettavat oikeudelliset normit tulee muuttaa kyllä tai ei -tyyppisiksi säännöiksi, jolloin niihin perustuva päättely on mahdollista toteuttaa koneellisesti. Haasteita aiheuttavat normit, jotka sisältävät avoimen käsitteen kuten ”pakottava syy”. Tällaisten avoimia käsitteitä sisältävien normien muuttaminen tietokonetta ohjaavaksi säännöksi osoittautuu yleensä haastavaksi tai mahdottomaksi kuten oikeusministeriön arviomuistiossa on todettu. Ks. oikeusministeriön arviomuistio 2020, s. 11.

hallintotoiminnan käyttökohteiksi on tunnistettu esimerkiksi tarkastukset, analyysit ja ennakoinnit.⁹³

Tekoälyratkaisuja on kokeiltu myös kunnallisella tasolla esimerkiksi, kun Espoon kaupungissa tehtiin vuonna 2017 kokeilu, jossa yhdistettiin Espoon kaupungin sosiaali- ja terveystalvelujen asiakkuusdataa ja varhaiskasvatuksen dataa tekoälyn analysoitavaksi. Kokeilulla pyrittiin selvittämään, onko tekoälyä mahdollista hyödyntää palvelujen kohdentamiseen ennaltaehkäisevästi. Tekoäly havaitsi sille annetusta datasta lähes 300 riskitekijää, jotka ennakoivat lastensuojelun asiakkuutta useamman tekijän osuessa samalle henkilölle.⁹⁴

Automaatio ei kata välttämättä koko päätöksentekoprosessia, vaan hallinnon organisaatiolla voi olla käytössä pohja-automaatiota, jossa sähköinen järjestelmä laatii vakiolausekkeet ja -perustelut sisältävän päätösehdotuksen olemassa olevien tietojen pohjalta. Tämän jälkeen päätöspohja ohjautuu manuaalisesti virkamiehen tarkistettavaksi tai täydennettäväksi, jolloin varsinainen hallintopäätös tehdään manuaalisesti. Lisäksi automaatiota on mahdollista hyödyntää päätöksentekoprosessin muissa vaiheissa, jolloin automaatiota käytännössä hyödynnetään päätöksenteon alkuvaiheen ohessa myös päätöksenteon keski- ja loppupuolen vaiheissa.⁹⁵

Tällä hetkellä käynnissä olevan hallinnon automaattista päätöksentekoa koskevan kansallisen lainsäädäntöhankkeen⁹⁶ työryhmämietinnön perusteella lainsäädännön muutoksen jälkeen hallinnon organisaatioiden olisi mahdollista hyödyntää hallintopäätösten automaatiossa ainoastaan sääntöperusteisia algoritmeja, eikä esimerkiksi koneoppimiseen perustuvia algoritmeja.⁹⁷

Tämän tutkielman laatimisen aikana oikeustila on kuitenkin epäselvä sen suhteen, miten pitkälle hallinnon organisaatioiden on mahdollista hyödyntää tekoälyä automaatiossa.

⁹³ Oikeusministeriön arviomuistio 2020, s. 13.

⁹⁴ 6Aika, Espoon tekoälykokeilun jatko: syrjäytymisriskin tekijöihin päästy kiinni, merkittäviä kustannussäästöjä jo nyt <<https://6aika.fi/espoon-tekoalykokeilun-jatko-syrjautymisriskin-tekijoihin-paasty-kiinni-merkittavia-kustannussaastoja-jo-nyt/>> Espoo>Uutiset>Asiakkuus 23.10.2019 (haettu 17.3.2022).

⁹⁵ Oikeusministeriön arviomuistio 2020, s. 18.

⁹⁶ Oikeusministeriön 18.1.2021–31.12.2021 väliselle ajalle asettaman työryhmän tehtävänä oli valmistella hallituksen esityksen muodossa ehdotus hallinnon automaattista päätöksentekoa koskevaksi yleislainsäädännöksi. Ks. Oikeusministeriön työryhmämietintö 2022, s. 8.

⁹⁷ Hallintolakiin ehdotettavan 53 e §:n perusteella viranomainen voi ratkaista automaattisesti asian, johon ei sisälly seikkoja, jotka edellyttäisivät tapauskohtaista harkintaa. Automaattisesti ratkaistaviksi asiat tulee valita ja ne tulee myös ratkaista sovellettavan lain perusteella laadittavien sääntöjen mukaisesti. Automaattisesti ei ole sallittua ratkaista oikaisuvaatimusta tai sitä vastaavaa oikeussuojakeinoa koskevaa asiaa, hallintokanteluasiaa tai rangaistusluonteisen hallinnollisen seuraamuksen määräämistä koskevaa asiaa. Ks. Oikeusministeriön työryhmämietintö 2022, s. 98. Vaatimus asian ratkaisemisesta kuvaamisesta nimenomaan sääntönä, sulkee pois sellaiset tekniset toteutustavat, joiden perustana on sääntöjen sijasta esimerkiksi todennäköisyyksien arviointi (esimerkiksi oppivat tekoälyteknikat). Ks. Oikeusministeriön työryhmämietintö 2022, s. 73.

Lisäksi tutkielman löydökset palvelevat osaltaan aihepiiriä koskevaa keskustelua, vaikka esimerkiksi koneoppimiseen perustuva automaatio kiellettäisiin hallintopäätösten osalta. Erityisesti on huomattava, että koneoppimiseen perustuva automaatio voi syrjiä rekisteröityä myös tosiasiallisen hallintotoiminnan tai valmistelevan toiminnan yhteydessä, mikäli kyseeseen tulee rekisteröidyn luokittelu (profilointi) ja siinä ilmenevä syrjintä.⁹⁸

Tämän tutkielman kannalta olennaista ei lopulta olekaan se, miten kansallisella tasolla päädytään lopulta säätelemään tekoälyyn perustuvasta hallinnon päätöksenteosta, vaan olennaista on huomata, että tekoälyjärjestelmän luokittellessa ihmisiä tiettyihin luokkiin, korostuvat tekoälyn mahdollinen syrjivyyden ja tämän tunnistettavuuden haasteet.⁹⁹

Nähtäväksi jääkin, millaisessa muodossa hallinnon automaattista päätöksentekoa koskeva sääntely lopulta saa päivänvalon. Joka tapauksessa teknologinen kehitys ja globaali tekoälykilpailu eivät osoita hiipumisen merkkejä, jolloin tänään hahmotellut hallinnon automaation rajoitteet saattavat osoittautua tarpeettoman jäykiksi ja toimimattomiksi yllättävänkin nopealla aikajänteellä.¹⁰⁰

2.4. Algoritmisten päätöksentekojärjestelmien syrjivyyden

2.4.1. Johdanto algoritmisten päätöksentekojärjestelmien syrjivyyteen

Algoritmisen päätöksenteon hyödyntäminen voi johtaa erinäisten toiminnan tehostumisen mukanaan tuomien hyötyjen lisäksi myös kielteisiin seurauksiin. Euroopan unionin tekoälyn valkoisessa kirjassa tunnistetaan erilaisia algoritmiseen päätöksentekoon liittyviä riskejä. Tällaisia riskejä ovat muun muassa tekoälyä hyödyntävän automaattisen päätöksenteon aiheuttama läpinäkymätön päätöksenteko, syrjintä ja yksityisyyden suojan heikentyminen algoritmisen päätöksenteon seurauksena.¹⁰¹ Euroopan parlamentti taas on tunnistanut algoritmisen päätöksenteon aiheuttavan riskejä tasa-arvolle, yksityisyydensuojalle¹⁰²,

⁹⁸ Ks. profiloinnin syrjivyyteen liittyen tämän tutkielman alaluku 2.4.2. Tässä tutkielmassa ei käsitellä tarkemmin tosiasiallisen hallintotoiminnan ja valmistelun automaation osalta syrjivyyttä. Esimerkkinä oikeusvaikutuksellisesta valmistelutoimesta voidaan nostaa esille TE-toimiston antama työvoimapoliittinen lausunto, joka merkittävästi vaikuttaa Kelan tai työttömyyskassan tekemään hakijaa koskevaan etuuspäätökseen, jolloin työvoimapoliittisesta lausunnosta voi seurata asianosaiselle merkittäviä oikeusvaikutuksia. Ks. Oikeuskanslerin lausunto, 19.8.2021. OKV/1698/21/2021, s. 4–5.

⁹⁹ Hakkarainen ym. 2020, s. 22,43. Ks. myös alaluku 1.1. tekoälyn syrjivyydestä ja syrjivyyden tunnistamisen haasteista.

¹⁰⁰ Ks. eurooppalaisesta lähestymistavasta tekoälyyn Euroopan komissio, A European approach to artificial intelligence <<https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>> Home>Policies>A European approach to artificial intelligence (haettu 21.3.2022).

¹⁰¹ Euroopan komission valkoinen kirja tekoälystä 2020, s. 1.

¹⁰² Algoritmisen päätöksenteko voi luoda riskejä yksityisyyden suojan toteutumiselle usealla tavalla. Ensinnäkin algoritmisen päätöksenteko saattaa edellyttää suurta määrää henkilötietoja, joka taas saattaa johtaa arvaamattomiin seurauksiin yksittäisille henkilöille. Algoritmien koulutukseen tarvitaan tyypillisesti suuria määriä henkilötietoja, mikä taas osaltaan lisää tietomurron riskiä ja sitä, että yksittäisen henkilön henkilötiedot joutuvat väärin käsiin. Ihmiset saattavat myös muuttaa käytöstään, mikäli heillä herää epäily siitä, että heidän tietojensa käytetään algoritmiseen päätöksentekoon. Ks. Euroopan parlamentin tutkimus 2019, s. 12–13.

ihmisarvolle, autonomialle, vapaalle tahdolle sekä terveydelle, elämänlaadulle, hyvinvoinnille ja fyysiselle koskemattomuudelle.¹⁰³

Voidaan siis todeta, että algoritmiseen päätöksentekoon saattaa sisältyä merkittäviä riskejä, jotka vaikuttavat yhteiskuntatason lisäksi yksilötasolla. Tässä tutkielmassa keskitytään nimenomaan yksilötasolla realisoituvaan syrjintään ja syrjivyyden tunnistettavuuteen.¹⁰⁴ Algoritmisen päätöksenteon syrjivyyden on monisyinen asia, eikä se kaikissa tapauksissa suinkaan ole järjestelmän kehittäjän tarkoitus. Voidaankin sanoa, että algoritmisen päätöksenteko voi johtaa henkilöiden syrjintään joko tarkoituksenmukaisesti tai tahattomasti.¹⁰⁵

Algoritmisen päätöksenteko perustuu erilaisiin luokittelutehtäviin. Binäärinen luokittelutehtävä on esimerkiksi päätös siitä, myönnetäänkö henkilölle laina vai ei. Moniluokkainen luokittelutehtävä taas on esimerkiksi sen määrittäminen, mihin etuusohjelmaan henkilö määritellään kuuluvaksi. Regressiotehtävässä taas määritellään esimerkiksi henkilölle tietty lainan korko. Syrjintää voi tapahtua vain, mikäli luokittelun tulokset on mahdollista asettaa paremmuusjärjestykseen toistensa kanssa (toisin sanoen, kyseeseen ei tule vaikutuksiltaan samanlainen seuraus riippumatta luokittelun tuloksesta).¹⁰⁶

Algoritmisen päätöksenteon syrjivyyden voi saada alkunsa jo järjestelmän kehitysvaiheessa järjestelmään päätyneiden vinoumien takia.¹⁰⁷ Tarkemmin sanottuna algoritmisen päätöksenteon syrjivyyden aiheuttajina voi olla erilaisia tekijöitä kuten järjestelmän opetusdatan¹⁰⁸ virheet ja puutteet, malliin valitut väärät ominaisuudet tai se, että päätöksentekosäännöt antavat merkitystä syrjintäperusteelle.¹⁰⁹

Edellä kuvatun perusteella havaitaan, että algoritmit itsessään eivät voi syrjiä, vaan tämän aiheuttavat muut algoritmiin liittyvät tekijät.¹¹⁰

¹⁰³ Euroopan parlamentin tutkimus 2019, s. 7.

¹⁰⁴ On kuitenkin huomattava, että syrjintä ilmiönä koskettaa myös laajemmin ihmisryhmiä, mutta tässä tutkielmassa asiaa tarkastellaan yksilön näkökulmasta. Esimerkiksi tässä tutkielmassa käsiteltävät tietosuojalainsäädännön mukaiset oikeussuojakeinot keskittyvät yksilön oikeuksiin johtuen henkilötietojen suojan perusoikeuslähtökohdista, mutta algoritmien haitat ilmenevät tyypillisesti ryhmätason syrjintänä. Ks. Edwards-Veale 2017, s. 22.).

¹⁰⁵ Yhdenvertaisuusvaltuutettu, tekoäly ja yhdenvertaisuus. <<https://syrjinta.fi/tekoaly#syrjintakiellon-valvonta>> Etusivu>Syrjintä ja yhdenvertaisuus>Aiheet>Tekoäly (haettu 21.3.2022).

¹⁰⁶ Zliobaite 2015, s. 6

¹⁰⁷ Euroopan parlamentin tutkimus 2019, s. 7.

¹⁰⁸ Opetusdata kouluttaa mallia käyttäytymään tietyllä tavalla. Tällöin opetusdatan luonteella voi olla merkitystä siihen, mitä järjestelmä oppii (ks. Barocas –Selbst 2016, s. 680.).

¹⁰⁹ Yhdenvertaisuusvaltuutetun lausunto, arviomuistio hallinnon automaattiseen päätöksentekoon liittyvistä yleislainsäädännön sääntelytarpeista VVTDno-2020–620, Asia: VN/3071/2020, s. 2.

¹¹⁰ Zliobaite 2015, s. 7.

Syrjivyyttä aiheuttaa erityisesti se, että tietyt vähemmistöt ovat historian kuluessa päätyneet aliedustetuiksi tai heitä kohtaan on havaittavissa monimutkaisempia tai vähemmän selvitettyjä kaavamaisuuksia verrattuina muihin. Tämän seurauksena mallit voivat suoriutua näitä ryhmiä koskevista päätöksistä erilaisella suorituskyvillä. Datalla toimiviin järjestelmiin vaikuttaa tietysti myös se, että käytännössä ei ole mahdollista kerätä kaikkea dataa kaikista yksilöistä sisältäen kaikki relevantit tietyn sosiaalisen ilmiön kannalta relevantit tiedot.¹¹¹

Algoritmien syrjivyydestä keskusteltaessa on syytä kiinnittää huomiota erityyppisiin algoritmeihin. Parametrisissa algoritmeissa (”parametric algorithm”) esiintyy yleensä enemmän vinoumia, mikä toisaalta mahdollistaa niiden nopeasti tapahtuvan asioiden oppimisen. Parametriset algoritmit eivät kuitenkaan ole erityisen tarkkoja ennustuksissaan monimutkaisten ongelmien yhteydessä. Muita tyypillisiä algoritmeja ovat päätöspuut, jotka sisältävät vähemmän vinoumia ja lineaariseen regressioon perustuvat algoritmit, jotka taas sisältävät runsaasti vinoumia tyypillisesti.¹¹²

Eryteisesti ihmisten luokitteluun hyödynnettävät algoritmiset järjestelmät voivat olla syrjiviä. Tällaisten järjestelmien tekemät päätökset nimittäin perustuvat siihen, että tietty yksilö määritellään kuuluvaksi tiettyyn luokkaan perustuen kyseisen yksilön ominaisuuksiin. Kun yksilöt on luokiteltu tiettyihin luokkiin kuuluviksi, voidaan tämän pohjalta toteuttaa päätöksiä myöhemmässä vaiheessa. Tässä yhteydessä olennaisiksi tulevat riskit luokittelussa tapahtuvista vääristä positiivisista ja vääristä negatiivisista ja näiden aiheuttamat seuraukset.¹¹³ On kuitenkin huomattava, että kaavamaisuuksien perusteella toimiviin koneoppimisjärjestelmiin liittyy sisäänrakennettuna ominaisuutena tietynlaista syrjivyyttä, sillä niiden tarkoituksena on erotella tietopisteitä toisistaan.¹¹⁴

Maailmalta on saatavilla esimerkkejä hallinnon organisaatioiden käyttämistä luokitteluun perustuvista algoritmista päätöksentekojärjestelmistä, joiden yhteydessä on havaittu syrjivyyttä. Niin sanotussa COMPAS-tapauksessa oli kyse siitä, että muun muassa New Yorkin, Wisconsinin ja Floridan osavaltioiden tuomioistuinten hyödyntämä päätöksenteon tukijärjestelmä COMPAS määrittäi säännönmukaisella tavalla rikoksen uusimisriskin matalammaksi valkoisille tuomituille verrattuna muuhun väestöön. COMPAS-tapaus nosti

¹¹¹ Veale –Binns 2017, s. 2.

¹¹² Ks. [Machine Learning Algorithms Mini-Course](https://machinelearningmastery.com/machine-learning-algorithms-mini-course/) (haettu 21.3.2022).

¹¹³ Diakopoulos 2016, s. 57.

¹¹⁴ Veale – Binns 2017, s. 2.

esille Hakkaraisen ym. mukaan myös tässäkin tutkielmassa käsiteltävän yksilön oikeusturvatakeita koskevan kysymyksen, eli sen, miten voidaan varmistaa, että rekisteröidylle annetaan riittävät ja oikeat tiedot päätöksenteosta.¹¹⁵

Kotimainen esimerkki algoritmisen päätöksenteon syrjivästä vaikutuksesta on yhdenvertaisuus- ja tasa-arvolautakunnan päätös koskien luottokelpoisuuden arviointimenettelyä. Kyseisessä tapauksessa oli kyse siitä, että luottolaitosyhtiö oli evännyt hakijalta laskuluoton myöntämisen hakijan asuinpaikan ja tämän sukupuolen, iän ja kielen perusteella. Yhtiö ei ollut huomionnut hakijaa koskevia tietoja tämän omasta luottokäyttäytymisestä ja -kelpoisuudesta. Hakijan luottokelpoisuuden arviointi ei ollut perustunut hakijan luottokelpoisuuden yksilölliseen arviointiin, vaan käytössä oli tilastollinen arviointimenetelmä, joka päätöksen mukaan olennaisin osin perustui yhdenvertaisuuslain (1325/2014) 8 §:n ja naisten ja miesten välisestä tasa-arvosta annetun lain (609/1986) jäljempänä ”tasa-arvolaki”) perusteella kiellettyihin syrjintäperusteisiin. Yhdenvertaisuus- ja tasa-arvolautakunnan mukaan luottolaitosyhtiö ei ollut pystynyt kumoamaan asiassa esitetyn perusteella syntyneitä syrjintäolettamia, jolloin yhtiön menettely hakijaa kohtaan oli yhdenvertaisuuslaissa ja tasa-arvolaisissa kiellettyä välitöntä moniperusteista syrjintää tämän sukupuoleen, äidinkieleen, ikään ja asuinpaikkaan liittyvien syiden johdosta.¹¹⁶

Myöskään sinänsä paikkansapitävän datan hyödyntäminen ei ole kaikissa tilanteissa ongelmaton. Barret nostaa esimerkiksi tilanteen, jossa algoritmin hyödyntämä data voi kertoa esimerkiksi tiedon siitä, että tietty määrä vähemmistöön kuuluvia henkilöitä on ollut vangittuna kuluneen vuoden aikana ja että tietty määrä tullaan vangitsemaan heidän elinaikanaan, mikäli samanlainen kehityskulku jatkuu. Barretin mukaan kyseisen tiedon hyödyntäminen ennustettaessa yksilön todennäköisyyttä syyllistyä rikokseen olisi moraalitonta, epäeettistä ja todennäköisesti perustuslain vastaista. Hänen mukaansa kyseessä olisi yksilön toimijuuden sekoittaminen väestöllisiin todennäköisyyksiin ja hämyisiin korrelaatioihin.¹¹⁷

¹¹⁵ Hakkarainen ym. 2020, s. 16.

¹¹⁶ Yhdenvertaisuus- ja tasa-arvolautakunnan päätös, 21.3.2018, 216/2017.

¹¹⁷ Barrett 2017, s. 340. Myös edellä mainittua yhdenvertaisuus- ja tasa-arvolautakunnan luottoyhtiötä koskevaa päätöstä voidaan arvioida tältä kannalta. Ratkaisussaan se nimittäin katsoi, että luottoyhtiön käyttämä pisteytysarviointi merkitsi sitä, että muita kuin luotonhakijaa koskevista tilastollisista tiedoista ja maksuhäiriöistä oli tehty oletuksia luotonhakijan luottokelpoisuudesta. Ks. Yhdenvertaisuus- ja tasa-arvolautakunnan päätös, 21.3.2018, 216/2017. Ks. myös Hirvonen 2018, s. 307.

Ihmisten toteuttama syrjintä tapahtuu tyypillisesti tapauskohtaisesti, kun taas algoritminen syrjintä voi olla systemaattista ja voi tapahtua suuremmissa mittaluokissa.¹¹⁸

Tekoälyn hyödyntämistä koskevassa keskustelussa on tarpeen huomata myös tekoälyn mahdollisuudet edistää yhdenvertaisuutta yhteiskunnassa.¹¹⁹ On nimittäin esitetty näkemyksiä, että automaattiset järjestelmät arvioivat kaikki yksilöt yhdenmukaisella tavalla siten välttämällä syrjinnän tapahtumisen. Kuten aiemmin on kuvattu, on kuitenkin mahdollista, että ihmisten vinoumat ja arvot ovat ikään kuin upotettuna järjestelmän kehitysvaiheeseen ja dataan. Pasqualen mukaan voidaankin todeta, että tietokoneistuminen saattaa siirtää syrjinnän ylävirtaan.¹²⁰

Seuraavaksi tarkastellaan algoritmisen päätöksenteon syrjivyyden seurauksia yksilöiden kannalta tarkasteltuna. Tämä luo kontekstia myös yksilöiden tiedollisten tarpeiden täyttämistä koskevaan keskusteluun tutkielmassa.

2.4.2. Algoritmisen päätöksenteon syrjivyyden seuraukset

Algoritmisen päätöksenteon syrjivyydestä keskustelu edellyttää myös tarkempaa analyysiä siitä, millaisia kielteisiä seurauksia syrjinnästä voi yksilölle aiheutua. Esimerkkinä tällaisista seurauksista voidaan mainita luottoluokitukset, jotka perustuvat algoritmiseen päätöksentekoon. Yhdysvalloissa tehdyn tutkimuksen perusteella vakuutuksessa käytettävät pisteet ovat huomattavasti huonommat alueilla, joissa asuu runsaasti vähemmistöjä. Luottoluokitusten osalta on havaittavissa kielteisiä vaikutuksia esimerkiksi työsuhteen, vakuutuksen tai vuokra-asuntojen saatavuuden yhteydessä.¹²¹

Suojeltujen ryhmien syrjintä näyttäytyy pisteytysjärjestelmien osalta siten, että näille ryhmille määritellään korkeampi riskikerroin tai muita haitallisia attribuutteja suuremmalla kertoimella kuin muille ryhmille.¹²² Maineseen perustuvat järjestelmät voivat luoda uusia ja useimmiten näkymättömiä vähemmistöjä, joita kohdellaan kielteisellä tavalla.¹²³ Myös profilointi voi vaikuttaa kielteisellä tavalla henkilön vapautteen valita esimerkiksi tiettyjä tuotteita tai palveluja. Lisäksi joissain tapauksissa profiloinnin perusteella voi syntyä epätarkkoja ennusteita. Profiloinnin tuloksena henkilöltä voidaan myös evätä palveluita ja

¹¹⁸ Zliobaite 2015, s. 4.

¹¹⁹ Yhdenvertaisuusvaltuutettu, tekoäly ja yhdenvertaisuus. <<https://syrjinta.fi/tekoaly#syrjintakiellon-valvonta>> Etusivu>Syrjintä ja yhdenvertaisuus>Aiheet>Tekoäly (haettu 21.3.2022).

¹²⁰ Pasquale 2015, s. 35.

¹²¹ Euroopan parlamentin tutkimus 2019, s. 8.

¹²² Zarsky 2014, s. 1384.

¹²³ Pasquale 2015, s. 38.

tavaroita ja tähän voidaan kohdistaa muuta perusteetonta syrjintää.¹²⁴ On tärkeää tunnistaa myös algoritmisen päätöksenteon ja profiloinnin vaikutukset yksittäisten henkilöiden mahdollisuuksiin päästä julkisten palveluiden piiriin yhdenvertaisesti.¹²⁵

Henkilötietoja koskevat algoritmiset analyysit voivat myös aiheuttaa erilaisia vaurauden siirtymisiä eri sosiaaliluokkien välillä, joita ei voida pitää hyväksyttävänä.¹²⁶ Syrjinnällä voi olla myös turvallisuuteen liittyviä vaikutuksia, mikäli algoritmisen päätöksenteon perusteella esimerkiksi virkavalta alkaa valvoa jotakin aluetta enemmän kuin toista. Lisäksi syrjivä algoritmisen päätöksenteko voi vahvistaa vähemmistöjä koskevia ennakkoluuloja.¹²⁷

Myös tietosuoja-asetuksen johdanto-osan 75 kappaleessa tunnistetaan, että henkilötietojen käsittelystä voi aiheutua rekisteröidylle fyysistä, aineellista tai aineetonta vahinkoa käsittelyn johtaessa rekisteröidyn syrjintään.¹²⁸

Hallinnon kontekstissa algoritmisen päätöksenteon syrjivyyden seuraukset riippuvat siitä, miten laajasti hallinnossa ylipäänsä otetaan algoritmista päätöksentekoa käyttöön. Huomioiden hallinnon toimintakentän laajuuden (hallintopäätökset, tosiasiallinen hallintotoiminta), edellä kuvatun perusteella voidaan todeta, että potentiaaliset syrjivän algoritmisen päätöksenteon ja automaation seuraukset voivat kuitenkin olla vakavia ja kauaskantoisia.¹²⁹

¹²⁴ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 6.

¹²⁵ Ks. personoinnin vaikutuksista eri palveluihin Hildebrandt – Gutwirth 2008, s. 121.

¹²⁶ Zarsky 2016, s. 123.

¹²⁷ Euroopan parlamentin tutkimus 2019, s. 9.

¹²⁸ Tietosuoja-asetuksen johdanto-osan 75 kappaleessa mainitaan seurauksina syrjinnän lisäksi tässä yhteydessä muun muassa identiteettivarkaus, petos, taloudelliset menetykset ja maineen vahingoittuminen.

¹²⁹ Ks. algoritmisen päätöksenteon käyttöalasta hallinnossa edellä kuvattu hallintolain automaattisen päätöksenteon muutosta koskeva lainsäädäntövalmistelutyö (erityisesti Oikeusministeriön työryhmämietintö 2022, s. 57–59).

3. TAUSTALLA VAIKUTTAVA LAINSÄÄDÄNTÖ

3.1. Perusoikeudet

Hallinnon organisaatioiden ottaessa käyttöön algoritmisia päätöksentekojärjestelmiä, tulee niiden varmistua järjestelmien lainmukaisuudesta mukaan lukien siitä, että automatisoidulle julkisen vallan käyttämiselle on löydettävissä peruste lainsäädännöstä. Edellä mainittu vaatimus korostuu tehtäessä hallintopäätöksiä, joilla on vaikutusta yksilöihin. Algoritmisia päätöksentekojärjestelmiä hyödyntävän organisaation on varmistuttava, että järjestelmä tuottaa vain lakiin perustuvia päätöksiä, ja että päätöksenteko noudattaa tarkoin lakia. Tämä tarkoittaa käytännössä sitä, että järjestelmän tulee noudattaa tarkoin päätösten perusteen muodostavaa aineellista lainsäädäntöä, mutta myös sitä, että hallinnon yleislakien kuten hallintolain, säännösten tulee toteutua.¹³⁰

Seuraavaksi kuvataan tarkemmin algoritmisen päätöksenteon syrjivyyteen liittyviä normeja hallinnon kontekstissa painottuen erityisesti yhdenvertaisuutta ja tasa-arvoa edistävään normistoon ja henkilötietojen käsittelyyn liittyvään normistoon.

Perusoikeudet muodostavat olennaisen osan hallintotoiminnan lainmukaisuutta, mikä merkitsee sitä, että viranomaisen on toiminnassaan arvioitava erilaisia ratkaisu- ja toimenpidevaihtoehtoja perusoikeussuojan toteutuminen huomioiden. Viranomaisen tulee arvioida muun muassa lakiin perustuvien rajoitusten oikeasuhtaisuutta, mahdollisten velvoitteiden täsmällisyyttä ja tarkkarajaisuutta sekä velvoitteiden toteuttamisen kannalta hyväksyttäviä perusteita.¹³¹

Eryityisesti vapausoikeudet ovat relevantteja arvioitaessa viranomaisen hyödyntämän algoritmisen päätöksentekojärjestelmän syrjivyyteen liittyviä perusoikeuskysymyksiä. Vapausoikeuksilla viitataan klassisiin perusoikeuksiin, joiden olennainen tarkoitus on suojata yksilöä julkiseen valtaan ja hallintoon nähden. Vapausoikeudet asettavat siten rajoituksia ensisijaisesti hallinnolliselle sääntelylle ja puuttumiselle, mutta niiden voidaan katsoa myös velvoittavan ja suuntavan viranomaisten toimintaa. Kyseeseen tulee yksilön vapauspiirin turvaaminen viranomaisten puuttumiselta ja hallinnollisilta rajoituksilta.

¹³⁰ Oikeusministeriön arviomuistio 2020, s. 26. Tässä tutkielmassa ei analysoida hallinnon algoritmisen päätöksenteon suunnitteluvaiheen toimenpiteitä, joilla voidaan pyrkiä varmistumaan järjestelmien syrjimättömyydestä. Kroll ym. esittävät, että läpinäkyvyys ja jälkikäteinen auditointi ovat vain rajallisesti hyödyllisiä estämään yksittäisiä haitallisia lopputulemia ja siksi tällaisia tekniikoita tulisikin hyödyntää yhdessä algoritmin suunnitteluvaiheessa käytettävien tekniikoiden kanssa. Ks. Kroll ym. 2017, s. 682.

¹³¹ Mäenpää, hallinto-oikeus, sähköinen teos, II yleinen hallinto-oikeus, 1. julkishallinnon valtiosääntöiset perusteet, perus- ja ihmisoikeudet hallintotoiminnassa, perus- ja ihmisoikeuksien toteuttaminen hallintotoiminnassa, perustuslakiin sisältyvät perusoikeudet.

Vapausoikeuksien alle kuuluvia perusoikeuksia ovat muun muassa oikeus elämään sekä henkilökohtaiseen vapauteen ja turvallisuuteen sekä sananvapaus ja yksityiselämän suoja. Lisäksi Euroopan ihmisoikeussopimus (63/1999) sisältää vastaavan sisältöisiä vapausoikeuksia.¹³²

Tämän tutkielman kannalta olennainen perusoikeus on oikeus yhdenvertaiseen kohteluun. On kuitenkin syytä tässä kohtaa nostaa esille myös toinen liitännäinen perusoikeus, eli henkilötietojen suoja.¹³³ Henkilötietojen suojan merkitys perusoikeutena saa merkitystä algoritmisen päätöksenteon yhteydessä silloin, kun päätöksentekoon hyödynnetään henkilötietoja. Seuraavaksi käsitellään tarkemmin Suomen perustuslain (731/1999) takaamaa oikeutta yhdenvertaiseen kohteluun ja yksityiselämän suojaan.

3.1.1. Oikeus yhdenvertaiseen kohteluun

Yhdenvertaisuus ja syrjintäkielto saavat perustansa kansainvälisistä sääntelyinstrumenteista. Suomessa inkorporoidun Euroopan ihmisoikeussopimuksen 1 osan 14 artiklassa asetetaan syrjinnän kieltö. Euroopan unionista tehdyn sopimuksen (SEU)¹³⁴ 2 artiklassa vahvistetaan syrjimättömyyden merkitys unionille. Euroopan unionin perusoikeuskirjan¹³⁵ 20 artiklan mukaan kaikki ihmiset ovat yhdenvertaisia lain edessä. Perusoikeuskirjan 21 artiklassa asetetaan syrjintäkielto. Perusoikeuskirjan 23 artiklassa säädetään naisten ja miesten välisestä tasa-arvosta.¹³⁶

Perustuslain 6 §:ssä turvataan oikeus yhdenvertaiseen kohteluun, mistä seuraa, ettei ketään saa ilman hyväksyttävää perustetta asettaa eri asemaan sukupuolen, iän, alkuperän, kielen, uskonnon, vakaumuksen, mielipiteen, terveydentilan, vammaisuuden tai muun henkilöön liittyvän syyn perusteella.

Tämän tutkielman kannalta olennaista on havaita, että perustuslain yhdenvertaisuussäännös ei rajoitu ainoastaan tuomioistuimiin tai lakia soveltaviin viranomaisiin, vaan sen piiriin kuuluvat myös pääsääntöisesti julkisen vallan käyttö ja julkinen toiminta. Tämä johtuu siitä, että perustuslain 2 §:n 3 momentissa säädetyn oikeusvaltioperiaatteen mukaisesti julkisen vallan käytön tulee perustua lakiin ja kaikessa julkisessa toiminnassa tulee noudattaa tarkoin

¹³² Mäenpää, hallinto-oikeus, sähköinen teos, II yleinen hallinto-oikeus, 1. julkishallinnon valtiosääntöiset perusteet, perus- ja ihmisoikeudet hallintotoiminnassa, perus- ja ihmisoikeuksien toteuttaminen hallintotoiminnassa, vapausoikeudet ja hallintotoiminta.

¹³³ Ks. Korpisaari ym. 2018, s. 34–35.

¹³⁴ Virallinen lehti nro C 326, 26/10/2012 s. 0001–0390.

¹³⁵ 2012/C 326/02.

¹³⁶ Perusoikeuskirja on oikeudellisesti sitova instrumentti, joka koskee sen 51(1) artiklan perusteella unionin eri toimielinten, elinten ja laitosten lisäksi myös unionin jäsenvaltioita silloin, kun ne soveltavat EU-oikeutta. Ks. Oikeusministeriön arviomuistio 2020, s. 30.

lakia, johon lasketaan kuuluvaksi perustuslain 6 §:n yhdenvertaisuus- ja syrjäntäkieltosäännökset.¹³⁷

Yhdenvertaisuus on tapana ryhmitellä muodolliseen ja tosiasialliseen yhdenvertaisuuteen. Muodollisella (oikeudellisella) yhdenvertaisuudella viitataan siihen, että samanlaisia tapauksia tulee kohdella samalla tavalla. Esimerkkinä tästä on tilanne, jossa kaksi hakijaa täyttävät laissa säädetty julkinen etuuden myöntämisen edellytykset, jolloin heidän hakemuksensa tulee ratkaista samalla tavalla. Muodollinen yhdenvertaisuus on kiinteä osa jäljempänä tarkasteltavaa välittömän syrjinnän käsitettä, joka tarkoittaa ihmisten erilaista kohtelua perustuen johonkin ei-hyväksyttävään, eli kiellettyyn syrjäntäperusteeseen, kuten ihonväriin, sukupuoleen tai uskontoon.¹³⁸

Koneoppimisjärjestelmät pyrkivät tunnistamaan datasta kaavamaisuuksia ja erottelemaan tietopisteitä toisistaan, jolloin voidaan havaita, että koneoppimisjärjestelmässä erot ryhmien kesken saatavilla olevassa datassa voivat mahdollisesti johtaa eroavaisuuksiin mallien suorituksissa.¹³⁹

Luokittelussa käytettävät ominaisuudet voivat olla selkeästi yhdenvertaisuutta ja tasa-arvoa turvaavassa lainsäädännössä kuvattujen ns. kiellettyjen syrjäntäperusteiden mukaisia kuten henkilön sukupuoli tai alkuperä, mutta ne voivat olla myös ominaisuuksia kuten henkilön ajotyyli tai hänen ostokäyttäytymisensä. Jälkimmäisten kaltaisten ominaisuuksien perusteella tehdyissä päätöksissä herää kysymys siitä, onko tällaisten muuttujien huomioiminen asianmukaista.¹⁴⁰ Tässä tutkielmassa ei kuitenkaan arvioida tarkemmin rajanvetoa sen suhteen, mitä ominaisuuksia algoritmisessa päätöksenteossa on asianmukaista huomioida.¹⁴¹

Perustuslain 6 §:n 2 momentin syrjäntäkieltosäännös sisältää kiellon erilliskohtelusta, eli segregaatiossa. Esimerkkinä tästä on tilanne, jossa sinänsä yhtäläisiä palveluita tarjotaan erikseen eri väestöryhmille jonkin syrjäntäkiellon sisältämän perusteen mukaisesti. Tämä on

¹³⁷ Anttila – Ojanen 2019, s. 9.

¹³⁸ Anttila – Ojanen 2019, s. 8.

¹³⁹ Veale – Binns, 2017, s. 2.

¹⁴⁰ Ks. yhdenvertaisuuslain 8 §, jossa on lueteltu syrjinnän kiellon alaisia henkilöön liittyviä syitä. Ks. myös Edwards-Veale 2017, s. 30.

¹⁴¹ Päätöksenteon perustuminen vähämerkityksellisiin tai mitättömiin eroavaisuuksiin tai mielivaltaiselta vaikuttaviin tekijöihin herättää yhdenvertaisuuden ja tasa-arvon lisäksi kysymyksen päätöksenteon asianmukaisuudesta ja kohtuullisuudesta. Tässä yhteydessä on tarpeen nostaa esille tietosuoja-asetus, jonka yhtenä merkittävänä periaatteena on henkilötietojen käsittelyn kohtuullisuus (”fairness”). Henkilötietojen käsittelyn tuloksena aiheutuva epäoikeudenmukaisuus tulisi arvioitavaksi tietosuoja-asetuksen viitekehysessä siten asettaen omat vaatimuksensa algoritmisen päätöksenteon oikeudenmukaisuudelle. Kohtuullisuuden periaate edellyttää sitä, että rekisterinpitäjät huomioivat rekisteröidyn edut ja odotukset, eivätkä väärinkäytä tietoja. Periaatteen tarkoituksena on myös estää salassa tapahtuva tietojen käsittely ja varmistaa rekisteröidyn tietoisuus käsittelyn luonteesta ja tarkoituksesta. Ks. Korpisaari ym. 2018, s. 90.

kiellettyä, ellei kyseeseen tule jokin hyväksyttävä syy, jonka takia toimintaa voidaan pitää kyseisessä tilanteessa oikeutettuna.¹⁴²

Yhdenvertaisuuslaissa konkretisoidaan perustuslain, kansainvälisten ihmisoikeussopimusten ja EU-oikeuden sisältämät yhdenvertaisuus- ja syrjintäkieltonormit kansallisen lain tasolla.¹⁴³

Perustuslain 6 §:n 4 momentissa annetaan perustuslaillinen toimeksianto edistää sukupuolten välistä tasa-arvoa yhteiskunnallisessa toiminnassa ja työelämässä.¹⁴⁴ Tasa-arvolaisissa säädetään sukupuoleen perustuvan syrjinnän kiellosta ja sukupuolten välisen tasa-arvon edistämisestä. Sukupuoleen perustuvaa syrjintää ja tasa-arvon edistämistä koskevat asiat tulevat siten lähtökohtaisesti käsiteltäviksi tasa-arvolain, eikä yhdenvertaisuuslain alla.¹⁴⁵

Myös rikoslaki (39/1889) sisältää syrjintää koskevaa sääntelyä. Rikoslain 11 luvun 11 §:ssä säädetään syrjintärikoksen tunnusmerkistöstä, joka on soveltamisalaltaan yleinen, minkä seurauksena teko voi täyttää rikoslain mukaisen syrjinnän tunnusmerkistön, vaikka se ei täyttäisi vastaavia yhdenvertaisuuslain mukaisia syrjinnän edellytyksiä.¹⁴⁶

3.1.2. Oikeus yksityiselämän suojaan

Käsiteltäessä henkilötietoja osana algoritmista päätöksentekoa, korostuu henkilötietojen suojan merkitys. Henkilötietojen suojasta puhuttaessa on tarpeen tarkastella henkilötietojen suojan luonnetta perusoikeusjärjestelmän osana. Perustuslain 10 §:ssä säädetään yksityiselämän suojasta, jonka perusteella jokaisen yksityiselämä, kunnia ja kotirauha on turvattu, minkä lisäksi henkilötietojen suojasta säädetään tarkemmin lailla. Yksityiselämän suojan rinnalla ja sen osa-alueena, joskin osin laajempaan käsitteenä, on yleistynyt termi henkilötietojen suoja.¹⁴⁷

Myös Euroopan ihmisoikeussopimuksessa säädetään yksityiselämän suojasta. Euroopan ihmisoikeussopimuksen 8 artiklan mukaan jokaisella on oikeus nauttia yksityis- ja perhe-elämäänsä, kotiinsa ja kirjeenvaihtoonsa kohdistuvaa kunnioitusta. Lisäksi Euroopan

¹⁴² Anttila – Ojanen 2019, s. 16 ja Perustuslakivaliokunnan lausunto 37/2014 vp, hallituksen esitys eduskunnalle laeiksi perusopetuslain 13 §:n ja lukiolain 9 §:n muuttamisesta HE 136/2014 vp, s. 2.

¹⁴³ Anttila – Ojanen 2019, s. 65.

¹⁴⁴ Anttila – Ojanen 2019, s. 17.

¹⁴⁵ Anttila – Ojanen 2019, s. 77.

¹⁴⁶ Anttila – Ojanen 2019, s. 71–72.

¹⁴⁷ Korpisaari ym. 2018, s. 5–6.

unionin perusoikeuskirjassa säädetään yksityiselämän suojasta 7 artiklassa. Henkilötietojen suojasta on säädetty erikseen perusoikeuskirjan 8 artiklassa.

Henkilötietoja käsiteltäessä yksityisyyden suojan muodostavat lisäksi muut perusoikeudet, joita ovat oikeus kunniaan, oikeus yhdenvertaiseen kohteluun, oikeus henkilökohtaiseen koskemattomuuteen, oikeus ihmisarvoiseen kohteluun, oikeus turvallisuuteen sekä yhdenvertaisuus. Myös syrjinnän kieltä ja oikeus vaikuttaa itseään koskeviin asioihin muodostavat osan yksityisyyden suojasta henkilötietoja käsiteltäessä.¹⁴⁸

Koneoppimiseen ei liitykään pelkästään syrjivyyttä koskevia huolenaiheita, vaan myös kohtuullisuuden toteutumista koskevat riskit nousevat esille kuten yllä kuvattiin tietosuojasetuksen kontekstissa. Tässä yhteydessä on mahdollista pohtia esimerkiksi sitä, onko oikeudenmukaista päätellä henkilön soveltuvuus työtehtävään hänen verkossa tekemiensä hakujen perusteella tai onko henkilön taloudellinen tilanne ominaisuus, jota on asianmukaista käyttää luokittelussa.¹⁴⁹

3.2. Syrjintää koskeva kansallinen lainsäädäntö

Syrjintä ei ole käsitteenä yksiselitteinen. Ensinnäkin siihen voidaan lukea lainsäädännössä suojatut ominaisuudet ja piirteet kuten henkilön terveydentila, sukupuoli ja alkuperä. Näiden suojattujen ominaisuuksien ja piirteiden hyödyntäminen on lähtökohtaisesti kiellettyä päätöksenteossa. Asia ei kuitenkaan ole aivan näin yksiselitteinen: nimittäin joissakin tilanteissa voidaan johtaa oikeutus ottaa huomioon päätöksenteossa suojattuja ominaisuuksia ja piirteitä (esimerkiksi elinajanodotteen arvioimisessa sukupuolen huomioiminen). On kuitenkin huomattava, että yhteiskunnan arvot kehittyvät jatkuvasti ja sitä kautta myös aiemmin hyväksyttävänä pidetty erottelu piirteiden tai ominaisuuksien johdosta voi myöhemmin muuttua syrjiväksi. Lisäksi on huomattava, etteivät kaikki oikeudenmukaisuuden mittarit ole yleismaailmallisia.¹⁵⁰

Edellä kuvattua tematiikkaa sisältyy kansalliseen yhdenvertaisuuslakiin. Yhdenvertaisuuslaissa konkretisoidaan perustuslain, kansainvälisten ihmisoikeussopimusten ja EU-oikeuden sisältämät yhdenvertaisuus- ja syrjintäkieltonormit kansallisella tasolla.¹⁵¹

¹⁴⁸ HE 96/1998 vp, Hallituksen esitys eduskunnalle henkilötietolaiksi ja eräksi siihen liittyviksi laeiksi, s. 5.

¹⁴⁹ Edwards-Veale 2017, s. 30.

¹⁵⁰ Veale –Binns, 2017, s. 2.

¹⁵¹ Anttila – Ojanen 2019, s. 65.

Yhdenvertaisuuslain 1 §:n mukaan lain tarkoituksena on edistää yhdenvertaisuutta, ehkäistä syrjinnän tapahtumista ja tehostaa syrjinnän kohteeksi joutuneen oikeusturvaa. Yhdenvertaisuuden edistämiseksi tarkoitetaan tässä yhteydessä yhdenvertaisuuden toteutumista käytännössä sisältäen toimenpiteet, joilla pyritään varmistamaan ihmisten yhdenvertainen kohtelu ja parantamaan muihin nähden heikommassa asemassa olevien asemaa, jotta voidaan saavuttaa tosiasiallinen yhdenvertaisuus.¹⁵²

Yhdenvertaisuuslain voidaankin edellä mainittu huomioiden nähdä asettavan algoritmisia päätöksentekojärjestelmiä hyödyntäville viranomaisille konkreettisen velvoitteen varmistaa yhdenvertaisuuden toteutuminen ja syrjinnän ehkäiseminen järjestelmien osalta.¹⁵³

Yhdenvertaisuuslaissa on omaksuttu laaja käsite epäsuotuisan kohtelun osalta. Yhdenvertaisuuslain esitöiden mukaan epäsuotuisa kohtelu käsittää sellaisen menettelyn, jonka johdosta menettelyn kohteena oleva asetetaan muihin nähden huonompaan asemaan. Kyse voi olla esimerkiksi rajoitusten lisäksi jonkin yleisesti saatavilla olevan etuuden, palvelun tai oikeuden epäämisestä tai rajoittamisesta.¹⁵⁴ Viranomaisen on esimerkiksi harkinnanvaraista etuutta koskevaa päätöstä tehdessään noudatettava yhdenvertaisuuslain säännöksiä toiminnassaan.¹⁵⁵

Yhdenvertaisuuslain 8 §:ssä luetellaan erinäisiä henkilöön liittyviä syitä, joiden perusteella tapahtuva syrjintä on kiellettyä. Lain 8 §:n mukaan ketään ei ole sallittua syrjiä tämän iän, alkuperän, kansalaisuuden, kielen, uskonnon, vakaumuksen, mielipiteen, poliittisen toiminnan, ammattiyhdistystoiminnan, perhesuhteiden, terveydentilan, vammaisuuden, seksuaalisen suuntautumisen tai muun henkilöön liittyvän syyn takia.

Yhdenvertaisuuslaissa jaotellaan syrjintä joko välittömään tai välilliseen.¹⁵⁶ Yhdenvertaisuuslain 10 §:n mukaan kyseessä on välitön syrjintä, jos henkilöä kohdellaan henkilöön liittyvän syyn perusteella epäsuotuisammin¹⁵⁷ kuin toista henkilöä on kohdeltu,

¹⁵² Anttila – Ojanen 2019, s. 66–67.

¹⁵³ Ks. Oikeusministeriön arviomuistio 2020, s. 32.

¹⁵⁴ Ks. HE 19/2014 vp, Hallituksen esitys eduskunnalle yhdenvertaisuuslaiksi ja eräksi siihen liittyviksi laeiksi, sivu 70.

¹⁵⁵ Ks. HE 19/2014 vp, Hallituksen esitys eduskunnalle yhdenvertaisuuslaiksi ja eräksi siihen liittyviksi laeiksi, sivu 57.

¹⁵⁶ Ks. yhdenvertaisuuslain 10 § ja 13 §.

¹⁵⁷ Hallituksen esityksen mukaan epäsuotuisalla kohtelulla viitataan: ”sellaiseen menettelyyn, joko tekoon tai laiminlyöntiin, jolla menettelyn kohde asetetaan muihin nähden huonompaan asemaan. Kyse voi siten olla esimerkiksi rajoitusten, vaatimusten, rasitteiden tai velvoitteiden asettamisesta vain tietyille henkilöille tai henkilöryhmälle samoin kuin esimerkiksi jonkin yleisesti saatavilla olevan etuuden, palvelun tai oikeuden epäämisestä tai rajoittamisesta.” Ks. HE 19/2014 vp, Hallituksen esitys eduskunnalle yhdenvertaisuuslaiksi ja eräksi siihen liittyviksi laeiksi, s. 70.

kohdellaan tai kohdeltaisiin vertailukelpoisessa tilanteessa.¹⁵⁸ Syrjinnän kiellolle on kuitenkin eräitä poikkeuksia.¹⁵⁹

Yhdenvertaisuuslain 13 §:ssä kielletään välillinen syrjintä. Välillinen syrjintä merkitsee tilannetta, jossa näennäisesti yhdenvertainen sääntö, peruste tai käytäntö asettaa henkilön muita epäedullisempaan asemaan henkilöön liittyvän syyn perusteella. Syrjinnän kiellolle on myös eräitä poikkeuksia.¹⁶⁰

Tasa-arvolain tarkoituksena on sen 1 §:n mukaan estää sukupuoleen perustuva syrjintä ja edistää naisten ja miesten välistä tasa-arvoa sekä parantaa tämän tarkoituksen nojalla naisten asemaa erityisesti työelämässä. Lain tarkoituksena on myös estää sukupuoli-identiteetin tai sukupuolen ilmaisun perusteella tapahtuva syrjintä. Lain 7 §:ssä kielletään välitön ja välillinen syrjintä sukupuolen perusteella.

Viranomaisten tulee noudattaa toiminnassaan myös hallintolain mukaisia velvoitteita. Yhdenvertaisuuden ja syrjimättömyyden kannalta olennainen velvoite on hallintolain 6 §:n mukainen tasapuolisuuden vaatimus, jonka perusteella viranomaisen tulee kohdella samanlaisia tapauksia samalla tavoin ja eri tapauksia eri tavoin.¹⁶¹

Koneoppimisjärjestelmien oikeudenmukaisuutta arvioitaessa niin sanottu tilastollinen pariteetti (”statistical parity”) kuvastaa mielenkiintoisella tavalla pyrkimystä tasapuoliseen kohteluun. Tilastollinen pariteetti edellyttää, että yhtä suureen määrään jokaisesta ryhmästä sovellettaisiin samaa mahdollista lopputulosta.¹⁶² Tällainen ryhmiä koskeva

¹⁵⁸ Hallituksen esityksen mukaan ”syrjintää olisi muihin nähden epäsuotuisampi kohtelu vertailukelpoisessa tilanteessa. Vaatimus tilanteiden vertailukelpoisuudesta viittaa siihen yleiseen periaatteeseen, että ihmisiä on samanlaisissa tapauksissa kohdeltava samalla tavalla. Jotta tilanteet olisivat säännöksessä tarkoitettussa mielessä vertailukelpoisia, tulee niiden olla rinnastettavissa toisiinsa oikeudellisesti merkityksellisten seikkojen osalta.” Ks. HE 19/2014 vp, Hallituksen esitys eduskunnalle yhdenvertaisuuslaiksi ja eräksi siihen liittyviksi laeiksi, s. 70.

¹⁵⁹ Yhdenvertaisuuslain 11 §:n 1 momentin perusteella erilaisessa kohtelussa ei ole kyse syrjinnästä, jos kohtelu perustuu lakiin ja sillä muutoin on hyväksyttävä tavoite ja keinot tavoitteen saavuttamiseksi ovat oikeasuhtaisia. Yhdenvertaisuuslain 11 §:n 2 momentin mukaan erilainen kohtelu on kuitenkin sallittua myös silloin, jos kohtelun oikeuttamisperusteista ei ole säädetty, mikäli kohtelulle on perus- ja ihmisoikeuksien kannalta tunnistettavissa hyväksyttävä tavoite ja keinot tavoitteen saavuttamiseksi ovat oikeasuhtaisia. Yhdenvertaisuuslain 11 §:n 2 momentissa on eräitä poikkeuksia, jolloin edellä mainittua ei sovelleta.

¹⁶⁰ Yhdenvertaisuuslain 13 §:n mukaan kyseessä ei ole välillinen syrjintä, jos säännölle, perusteelle tai käytännölle voidaan tunnistaa hyväksyttävä tavoite ja tavoitteen saavuttamiseksi hyödynnetyt keinot ovat asianmukaisia ja tarpeellisia.

¹⁶¹ Oikeusministeriön arviomuistio 2020, s. 32.

¹⁶² Lepri ym. 2018, s. 6. Algoritmissen päätöksenteon syrjivyyden yhteydessä voidaan tarkastella myös algoritmisen päätöksenteon oikeudenmukaisuutta. Oikeudenmukaisuuden mittaamiselle on useita tapoja, joihin kuuluvat edellä kuvattu tilastollinen/demografinen pariteetti (”statistical/demographic parity”), jossa huomioidaan luokitteluasteita eri ryhmien kesken. Ennustusten oikeudenmukaisuus (”accuracy equity”) taas katsoo ennustavan mallin kokonaisvaltaista tarkkuutta jokaisen ryhmän osalta ja ehdollisen tarkkuuden oikeudenmukaisuus (”conditional accuracy equity”) taas kiinnittää huomiota ennustavan mallin tarkkuuteen jokaisen ryhmän osalta riippuen heidän ennustetusta luokastansa. Mahdollisuuksien oikeudenmukaisuudessa (”equality of opportunity”) arvioidaan sitä, soveltuuko jokaiseen ryhmään yhtä suurella todennäköisyydellä myönteinen lopputulos perustuen ryhmän perustasoon. Lisäksi toisenlaisessa väärinkohtelussa (”disparate mistreatment”) huomioidaan väärin positiivisten eroja eri ryhmien välillä. On olemassa myös erilaisia vaihtoehtoisia lopputuloksiin perustuvia tapoja arvioida oikeudenmukaisuutta kuten henkilön siirtämistä ei-suojeltuun

oikeudenmukaisuus ei kuitenkaan usein mahdollista tarkkaa oppimista ja tosiasiallista oikeudenmukaisuutta.¹⁶³ Hardt ym. ovat esittäneet oikeudenmukaisuuden mittaamistapaa, jossa pyritään huomioimaan tilastollisen pariteetin puutteet. Lisäksi siinä toteutetaan valvotun koneoppimisen kannalta tärkeä vaatimus tarkempien luokittelijoiden kehittämiseen. He kuitenkin tunnistavat haasteita esimerkiksi luotettavan datan saatavilla olemisen osalta.¹⁶⁴

Yllä kuvattu hallintolain 6 §:n tasapuolisuuden vaatimus on viranomaiskontekstissa tapahtuvan päätöksenteon reunaehto. On kuitenkin olennaista huomata, että viranomaisen hyödyntäessä syrjivää algoritmista päätöksentekojärjestelmää, saattaa kyseeseen tulla hallinto-oikeudellisten normien rikkomisen lisäksi yhdenvertaisuus- ja tasa-arvolainsäädännön vastainen menettely sekä tietosuoja-asetuksen vastainen henkilötietojen käsittely.¹⁶⁵

3.3. Tietosuoja-asetus ja automaattinen päätöksenteko

Algoritmiseen päätöksentekoon ei välttämättä sisälly lainkaan henkilötietojen käsittelyä (algoritmisen päätöksenteko voi koskea esimerkiksi yritysdataa, infrastruktuuria tai vaikkapa tehtaan tuotantoprosessia ilman tunnistettaviin henkilöihin liittyviä tietoja).¹⁶⁶ Tämän tutkielman kohteena on kuitenkin sellainen algoritmisen päätöksenteko, jossa päätöksiä tehdään yksittäistä henkilöä koskien hyödyntäen kyseistä henkilöä ja mahdollisesti muita henkilöitä koskevia henkilötietoja.

Tietosuoja-asetus sisältää yksityiskohtaista sääntelyä henkilöiden profilointia ja automatisoituja päätöksiä koskien. Tietosuoja-asetuksessa tunnistetaan näiden potentiaali

luokkaan sen arvioimiseksi, onko luokittelun lopputulos sama kuin jos hän kuuluisi suojeltuun luokkaan. Oikeudenmukainen järjestelmä täyttäisi kaikki edellä mainitut mittarit, mutta tutkimus on osoittanut, että on mahdotonta luoda mallia, joka täyttäisi nämä samanaikaisesti pois lukien poikkeukselliset tilanteet, jotka eivät kuitenkaan todennäköisesti olisi tosielämässä mahdollisia. Ks. Veale –Binns 2017, s. 3.

¹⁶³ Esimerkiksi lainojen yhteydessä algoritmin tarkkuus kärsii, mikäli sen tulee ennustaa yhtä suureksi kahden eri ryhmän maksukyvykyys, jos näiden välillä on todellisuudessa eroa. Ryhmiä koskeva oikeudenmukaisuus ei myöskään takaa sitä, että yhteen ryhmään kuuluva luottokelpoinen henkilö saisi esimerkiksi lainan yhtä todennäköisesti kuin toiseen ryhmään kuuluva samalla tavalla luottokelpoinen henkilö. Ks. Lepri ym. 2018, s. 6.

¹⁶⁴ Hardt ym. 2016, s. 19-21.

¹⁶⁵ Ks. esimerkiksi yhdenvertaisuuslain 3 ja 5 luvut, tasa-arvolain 7–12 §:t ja tietosuoja-asetuksen 5 artikla. Tässä tutkielmassa ei kuitenkaan arvioida sitä, millaisiin mahdollisiin seuraamuksiin syrjivä algoritmisen päätöksenteko johtaa hallinnon kontekstissa.

¹⁶⁶ Tietosuoja-asetuksen 4(1) artiklassa määritellään henkilötiedoksi kaikki tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön liittyvät tiedot. Tunnistettavissa olevana pidetään luonnollista henkilöä, joka voidaan suoraan tai epäsuorasti tunnistaa erityisesti tunnistetietojen, kuten nimen, henkilötunnuksen, sijaintitiedon, verkkotunnistetietojen taikka yhden tai useamman hänelle tunnusomaisen fyysisen, fysiologisen, geneettisen, psyykkisen, taloudellisen, kulttuurillisen tai sosiaalisen tekijän perusteella.

aiheuttaa riskejä yksityisyyden suojalle.¹⁶⁷ Tietosuoja-asetus myös asettaa rekisterinpitäjille velvoitteen teknisten ja organisatoristen suojakeinojen käyttöönosta.¹⁶⁸

Tietosuoja-asetuksessa on omaksuttu automaattista päätöksentekoa rajoittava lähestymistapa, sillä automaattinen päätöksenteko on käytännössä lähtökohtaisesti kielletty (tietosuoja-asetuksen 22(1) artiklan mukaan rekisteröidyllä on oikeus olla joutumatta tällaisen päätöksen kohteeksi). On kuitenkin tärkeää huomata, että tietosuoja-asetuksen erityiset automaattista päätöksentekoa koskevat reunaehdot ja velvoitteet koskevat nimenomaan asetuksessa määritetyn kaltaista päätöksentekoa.¹⁶⁹

Automaattinen päätöksenteko on kuitenkin sallittua tietyissä rajatuissa tapauksissa.¹⁷⁰ Profilointia ei ole kielletty tietosuoja-asetuksessa, mutta rekisterinpitäjien tulee noudattaa tietosuoja-asetuksen vaatimuksia kuten läpinäkyvyyttä ja asianmukaisten riskiarvioiden suorittamista käsittelystä.¹⁷¹

WP29 tietosuojatyöryhmän mukaan, mikäli automaattinen päätöksenteko on sallittu jonkin poikkeuksen nojalla, on rekisterinpitäjän toteutettava toimenpiteitä, joilla suojataan rekisteröidyn oikeudet ja vapaudet sekä oikeutetut edut mahdollistaen rekisteröidyn henkilötietojensa valvonta.¹⁷²

Toimenpiteitä ovat seuraavat:

- Käsittelystä ilmoittaminen rekisteröidylle.
- Rekisteröidyn oikeus vaatia, että ihminen osallistuu tietojen käsittelemiseen.
- Rekisteröidyn oikeus esittää kantansa, rekisteröidyn oikeus saada selvitys kyseisen arvioinnin jälkeen tehdystä päätöksestä ja rekisteröidyn oikeus riitauttaa päätös.¹⁷³

¹⁶⁷ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 6.

¹⁶⁸ Janssen 2020, s. 83.

¹⁶⁹ Automaattisen päätöksenteon määritelmä on kuvattu tietosuoja-asetuksen johdanto-osan 71 kappaleessa, jonka mukaan *''Rekisteröidyllä on oikeus olla joutumatta hänen henkilökohtaisia ominaisuuksiaan arvioivan, mahdollisesti toimenpiteen sisältävän päätöksen kohteeksi, joka on tehty yksinomaan automaattisen tietojenkäsittelyn perusteella, josta hänelle aiheutuu oikeudellisia vaikutuksia tai joka vaikuttaa häneen vastaavalla merkittäväällä tavalla, kuten online-luottihakemuksen automaattinen epäminen tai sähköisen rekrytoinnin käytännöt ilman ihmisen osallistumista.''* Ks. myös tietosuoja-asetuksen 22 artikla.

¹⁷⁰ Automaattinen päätöksenteko on sallittua, mikäli päätös on välttämätön rekisteröidyn ja rekisterinpitäjän välisen sopimuksen tekemistä tai täytäntöönpanoa varten; on hyväksytty rekisterinpitäjään sovellettavassa unionin oikeudessa tai jäsenvaltion lainsäädännössä; tai päätös perustuu rekisteröidyn nimenomaiseen suostumukseen. Ks. Tietosuoja-asetus 22(2) artikla.

¹⁷¹ Ks. profilointiin liittyen tietosuojavaltuutetun toimisto, automaattinen päätöksenteko ja profilointi <<https://tietosuoja.fi/automaattinen-paatoksenteko-profilointi>> Organisaatiot>Henkilötietojen käsittely>Automaattinen päätöksenteko ja profilointi (haettu 23.3.2022).

¹⁷² 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 21.

¹⁷³ Tietosuoja-asetuksen johdanto-osan 71 kappale ja tietosuoja-asetuksen 22(3) artikla. Tietosuoja-asetuksen 22(3) artiklassa ei tosin mainita kaikkia johdanto-osan 71 kappaleen sisältämiä oikeuksia kuten käsittelystä ilmoittamista, eikä oikeutta saada selvitystä kyseisen arvioinnin jälkeen tehdystä päätöksestä.

Toimenpiteisiin profiloinnin osalta taas voidaan katsoa kuuluvan seuraavat¹⁷⁴:

- Rekisterinpitäjän velvoite hyödyntää profiloinnissa asianmukaisia matemaattisia tai tilastollisia menetelmiä.
- Rekisterinpitäjän velvoite toteuttaa teknisiä ja organisatorisia toimenpiteitä, jotta henkilötietojen virheellisyyteen johtavat tekijät korjataan ja virheriski minimoidaan.
- Rekisterinpitäjän velvoite estää muun muassa luonnollisten henkilöiden syrjintä taikka vaikutukset, joiden johdosta toteutetaan toimenpiteitä, joilla on tällaisia seurauksia.¹⁷⁵

WP29 tietosuojatyöryhmä on laatinut ehdotuksia hyvistä käytännöistä rekisterinpitäjille niiden tehdessään tietosuoja-asetuksen 22(1) artiklassa määriteltyjä automaattisia päätöksiä sisältäen myös tilanteet, joissa mukana on profilointia. Käytännöt ovat osittain yhteneväisiä edellä kuvattujen toimenpiteiden kanssa ja täydentävät niitä. Alla on lueteltu tämän tutkielman kannalta olennaisimpia käytäntöjä:

- Järjestelmien säännölliset laadunvarmistustarkastukset sen varmistamiseksi, että henkilöitä kohdellaan oikeudenmukaisesti, eikä heitä syrjitä.
- Koneoppimisjärjestelmien hyödyntämien ja kehittämien algoritmien testaus, jotta voidaan osoittaa, etteivät ne tuota syrjiviä, virheellisiä tai perusteettomia tuloksia, vaan toimivat tarkoitetulla tavalla.
- Kolmansien osapuolten toimittamia algoritmeja käytettäessä varmistetaan sopimusteitse siitä, että algoritmi on tarkastettu, testattu ja vastaa sovitun standardin.
- Keinot, jotka mahdollistavat rekisteröidylle kantansa ilmaisun ja päätöksen vastustamisen.
- Ihmisen osallistumisen mahdollistava mekanismi tietyissä tilanteissa, esimerkiksi tarjoamalla linkki rekisteröidylle muutoksenhakuprosessiin automatisoidun päätöksen toimittamisen yhteydessä, uudelleentarkastelun määräajat ja nimetty yhteyspiste tiedusteluja varten.¹⁷⁶

Kansalliseen lainsäädäntöön sisältyvien viranomaisten toimintaa koskevien säännösten on mahdollista katsoa vastaavan tietosuoja-asetuksessa tarkoitettuja automaattisen päätöksenteon suoja-toimia. Tällaisiksi viranomaisten velvoitteiksi on tunnistettu esimerkiksi päätöksen perusteluvelvollisuus, velvollisuus antaa neuvontaa, velvollisuus hyvään

¹⁷⁴ Nämä suoja-toimet koskevat tietosuoja-asetuksen johdanto-osan 71 kappaleen mukaan vain profilointia, mutta tässä tutkielmassa painopiste on nimenomaan rekisteröityihin kohdennetussa profiloinnissa ja sen perusteella tehtävissä automaattisissa päätöksissä.

¹⁷⁵ Tietosuoja-asetuksen johdanto-osan 71 kappale.

¹⁷⁶ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 33–35.

kielenkäyttöön ja kuulemista koskevat säännökset. Olennainen suojatoimi on myös oikeus valittaa hallintopäätöksestä, eli oikeus riitauttaa päätös. Suojatoimia sisältyy myös viranomaisia koskevaan erityislainsäädäntöön esimerkiksi tilanteissa, joissa viranomaiset käsittelevät arkaluonteisia henkilötietoja. Lisäksi rekisteröidylle ilmoittaminen hallintoasian ratkaisemisesta automaattisesti, automaattisen päätöksentekojärjestelmän ja päätöksentekosääntöjen hyväksyminen, sekä automaattisen päätöksenteon valvonta voidaan katsoa suojatoimia vastaaviksi säädöksiksi.¹⁷⁷

¹⁷⁷ Oikeusministeriön arviomuistio 2020, s. 58.

4. ALGORITMISEN PÄÄTÖKSENTEON LÄPINÄKYVYYS JA YMMÄRRETTÄVYYS

4.1. Läpinäkyvyys laajasti ymmärrettynä

Algoritmisten päätöksentekojärjestelmien asianmukaisuus ja lainmukaisuus tulisi olla kenen tahansa todennettavissa ideaalitulanteessa. Etenkin hallinnon kontekstissa järjestelmän toiminnan kokonaisvaltainen ymmärtäminen, sekä sen testaaminen ja epäkohtien tunnistamisen mahdollisuus mahdollistaa sen, että kansalaiset voivat luottaa järjestelmään osana julkisen vallan käyttöä. Edellä kuvatun ideaalitulanteen saavuttaminen ei kuitenkaan ole yksinkertaista edes perinteisten hallintopäätösten, puhumattakaan algoritmisen päätöksentekojärjestelmän tuottamien päätösten osalta.¹⁷⁸

Vaatus algoritmisten järjestelmien läpinäkyvyyden lisäämisestä on ollut esillä akateemisessa keskustelussa jo useita vuosikymmeniä, mutta edelleenkin ei ole saavutettu yhteisymmärrystä siitä, millaisen muodon läpinäkyvyys tässä kontekstissa käsittäisi.¹⁷⁹

Läpinäkyvyys nousee esille EU:n valkoisessa kirjassa, jossa todetaan, että läpinäkyvyyden perusteella tulee antaa selkeä tieto tekoälyjärjestelmien kyvykkyyksistä ja rajoitteista: etenkin järjestelmien käyttötarkoitukset, toiminnallisuudet ja tarkkuus tulisi kuvata. Lisäksi EU:n valkoisessa kirjassa korostetaan sitä, että kansalaisille tulisi kertoa selkeällä tavalla heidän ollessaan tekemisissä tekoälyjärjestelmän kanssa. Tässä yhteydessä mainitaan erikseen EU:n tietosuojalainsäädäntö, mutta tämän lisäksi korostetaan mahdollisuutta muille vaatimuksille läpinäkyvyyden takaamiseksi. EU:n valkoisen kirjan mukaan on tärkeää, että informaatio annetaan objektiivisessa, tiiviissä ja helposti ymmärrettävässä muodossa tilanteen edellyttämän tarpeen mukaan.¹⁸⁰

Myös Suomessa lainsäätäjä on tiedostanut algoritmisen päätöksenteon läpinäkyvyyden merkityksen. Esimerkiksi hallituksen esityksessä eduskunnalle laiksi henkilötietojen käsittelystä maahanmuuttohallinnossa ja eräksi siihen liittyviksi laeiksi velvoitettiin maahanmuuttovirasto julkistamaan rekisteröidyille automaattisessa päätöksenteossa

¹⁷⁸ Hakkarainen ym. 2020, s. 22.

¹⁷⁹ Kroll ym. 2017, s. 638.

¹⁸⁰ Euroopan komission valkoinen kirja tekoälystä 2020, s. 20. Ks. Myös EU:n tekoälyasetusehdotuksen (ehdotus Euroopan parlamentin ja neuvoston asetus tekoälyä koskevista yhdenmukaistetuista säännöistä (tekoälysäädös) ja tiettyjen unionin säädösten muuttamisesta COM/2021/206 final) johdanto-osan 47 kappale, jonka mukaan suuririskisiltä tekoälyjärjestelmiltä olisi edellytettävä tiettyä läpinäkyvyyttä, jotta niiden läpinäkymättömyyteen voidaan puuttua. Läpinäkymättömyys voi tehdä tekoälyjärjestelmistä käsittämättömiä tai liian monimutkaisia ihmisten kannalta. Käyttäjien tulee olla mahdollista tulkita järjestelmän tuotosta ja hyödyntää sitä asianmukaisesti. Tästä johtuen suuririskisten tekoälyjärjestelmien mukana tulisi toimittaa asiakirjat ja käyttöohjeet sisältäen ytimekkäitä ja selkeitä tietoja (tarpeen mukaan myös mahdollisista perusoikeuksiin kohdistuvista riskeistä ja syrjintäriskeistä).

käyttämänsä lopulliseen päätökseen johtava algoritmi.¹⁸¹ Oikeuskansleri on ratkaisukäytännössään korostanut, että automaattisen päätöksenteon läpinäkyvyyden ja avoimuuden tarkoitus on korostaa ihmiskeskeistä ja ihmisen toimijuuteen perustuvaa hyvää hallintoa, mikä lisää luottamusta päätöksentekoa ja viranomaistoimintaa kohtaan. Luottamus taas vahvistaa automaation käytön oikeutusta ja hyväksyttävyyttä päätöksenteossa ja hallintotoiminnassa.¹⁸²

Riski päätöksen syrjivyydelle on suurin silloin, kun päätöksenteon perusteet eivät ole julkisia.¹⁸³ Hallinnon kontekstissa päätöksenteon sisällön tarkastelun mahdollistajina nousevatkin esille läpinäkyvyys ja avoimuus. Hakkaraisen ym. mukaan läpinäkyvyydellä viitataan laajaan julkisuuteen, omin silmin asioiden todentamiseen ja avoimuuteen. Heidän mukaansa avoimuus on mahdollista käsittää läpinäkyvyyden hallinto-oikeudelliseksi vastineeksi tai ainakin sille läheiseksi. Avoimuuden seurauksena yksilöiden on mahdollista valvoa valtion toimia.¹⁸⁴

Läpinäkyvyyden vastakohtana muodostaa läpinäkymättömyys, joka yhdistetään läpäisemättömyyteen ja tietynlaiseen harhaanjohtavuuteen. Läpinäkyvyys mielletään tilaksi, jossa ulkopuolisten on mahdollista seurata toimintaketjua ja päätöksentekoa, kun taas läpinäkymättömyys viittaa siihen, että tietty toiminta ja päätöksenteko on piilossa ulkopuolisten tarkastelulta ja arvioilta. Läpinäkymättömyys oletusarvoisesti haittaa tiedonhankintaa ja rationaalista keskustelua.¹⁸⁵

Edellä kuvatuissa läpinäkyvyyden hyödyissä on nähtävissä ensisijaisesti yleinen yhteiskunnallinen etu, mutta myös yksittäiset henkilöt hyötyvät julkisten organisaatioiden tuottamista päätöksentekoa koskevista selityksistä. He voivat esimerkiksi tietoisesti tutustua ja reagoida erilaisiin julkisiin hankkeisiin.¹⁸⁶ Asianosaisen oikeus saada perusteltu päätös luo edellytykset päätöksen oikeellisuuden kontrolloimiselle ja oikeussuojakeinojen käytölle.¹⁸⁷

Algoritmisille syönteille annettava läpinäkyvyys ja selitykset voivatkin palvella Diakopouloksen mukaan useita eri tavoitteita mukaan lukien ymmärrettävyyttä, luottamusta,

¹⁸¹ HE 18/2019 vp, Hallituksen esitys eduskunnalle laiksi henkilötietojen käsittelystä maahanmuuttohallinnossa ja eräiksi siihen liittyviksi laeiksi, s. 101.

¹⁸² Ks. Valtioneuvoston oikeuskansleri, oikeuskanslerin ratkaisu 2021, OKV/131/70/2020 20.4.2021.

¹⁸³ Oikeusministeriön arviomuistio 2020, s. 32.

¹⁸⁴ Hakkarainen ym. 2020, s. 8.

¹⁸⁵ Stohl ym. 2016, s. 124.

¹⁸⁶ Edwards-Veale 2017, s. 39.

¹⁸⁷ Hakkarainen ym. 2020, s. 18.

tehokkuutta, vakuuttavuutta, ja tyytyväisyyttä.¹⁸⁸ Zarskyn mukaan läpinäkyvyyden avulla on myös mahdollista korjata virheitä algoritmisessa prosessissa ja sitä kautta edistää prosessin tehokkuutta. Läpinäkyvyyden kautta yksilöiden on myös mahdollista korjata epätäsmällistä dataa, jota heistä on kerätty. Tätä kautta läpinäkyvyyden toteuttaminen myös luo painetta kehittää organisaatioiden toimintatapoja.¹⁸⁹

Algoritmisen päätöksenteon avoimuus tulee Hakkaraisen ym. mukaan asettaa julkisen päätöksenteon saavutettavuuden ja sitä kautta julkishallinnon asiakaslähtöisyyden kehittämisen kontekstiin, missä yhteydessä on kuitenkin huomattava, ettei läpinäkyvyys toteudu nykyisessäkään käytännössä kansalaisten suuntaan.¹⁹⁰

Hakkarainen ym. ovat tarkastelleet päätöksentekojärjestelmien avoimuuteen liittyviä kysymyksiä eri näkökulmista. Asiaa on mahdollista heidän mukaansa tarkastella perustuslaillisen ja hallinto-oikeudellisen julkisuusperiaatteeseen perustuvan järjestelmätason julkisuuden kautta. Tällöin kyseeseen tulisi avoimuuden toteuttaminen julkisuuslainsäädännön kautta algoritmisen järjestelmän lähdekoodin¹⁹¹ tai yksittäisen päätöksen tasolla. Toisaalta avoimuutta voidaan lähestyä myös päätösten perusteluvollisuuden kautta, jolloin siirrytään järjestelmätason julkisuudesta asianosaisavoimuuteen, eli rekisteröityjen kautta tarkasteltavaan avoimuuteen.¹⁹²

Avoimuus suhteessa algoritmisiin päätöksentekojärjestelmiin onkin Hakkaraisen ym. mielestä laaja käsite avoimuuden toteutustapojen kannalta, jolloin avoimuuden kohteita voivat olla esimerkiksi tietojärjestelmä, yksittäiseen päätöslajiin sovellettava automaatio, järjestelmän lähdekoodi tai vaikkapa järjestelmän toimintalogiikka.¹⁹³

Ribanan, Bohnin, Duarten ja Garcken mukaan koneoppimisalgoritmin läpinäkyvyys edellyttää, että mallin parametrejä opetusdatasta synnyttävät ja datan arvoja testidatasta luovat prosessit on mahdollista selittää ja perustella. Heidän mukaansa koneoppimisjärjestelmien läpinäkyvyys käsittää usean osa-alueen, joita ovat mallin kokonaisrakenne, yksittäiset mallin osat, oppimisalgoritmi sekä tieto siitä, miten tiettyyn

¹⁸⁸ Diakopoulos 2016, s. 61.

¹⁸⁹ Zarsky 2016, s. 122.

¹⁹⁰ Hakkarainen ym. 2020, s. 38.

¹⁹¹ Erityisesti lähdekoodin luonnetta viranomaisen asiakirjana on pohdittu akateemisessa keskustelussa muun muassa Hakkaraisen ym. toimesta peilaten epäselvään oikeustilaan tältä osin. (Ks. Hakkarainen ym. 2020, s. 25–28.) Tässä tutkielmassa ei arvioida tarkemmin sitä, voiko algoritmin opetusdata tai muu kuin sen koodi olla julkisuuslain alaista julkista tietoa. Joka tapauksessa julkisuuslain tehokkuus suhteessa päätöksenteon ymmärrettävyyteen on rajattu, sillä siinä ei ole erillistä velvoitetta selittää annettua tietoa. Tämän lisäksi viranomaisen asiakirjan tulee julkisuuslain 5 § 2 momentin mukaan olla viranomaisen hallussa, mikä tuonee omat haasteensa määritelmän täyttymiselle silloin, kun esimerkiksi opetusdataa ei ole luovutettu järjestelmää käyttävälle viranomaiselle (Ks. Hakkarainen ym. 2020, s. 26–27).

¹⁹² Hakkarainen ym. 2020, s. 4–5.

¹⁹³ Hakkarainen ym. 2020, s. 4.

ratkaisuun on päädytty algoritmin toimesta. He erottelevat myös mallin läpinäkyvyyden, suunnittelun läpinäkyvyyden ja algoritmien läpinäkyvyyden todeten kuitenkin, että yleisesti ottaen koneoppimisalgoritmien metodit eivät realistisesti voi olla täysin läpinäkyviä, vaan kyseeseen tulee ennemminkin eriasteinen läpinäkyvyys.¹⁹⁴

Läpinäkyvyyden toteuttaminen on mahdollista hahmottaa Anannyn ja Crawfordin mukaan eri tarpeita tarkastellen. Esimerkiksi joidenkin järjestelmien osalta läpinäkyvyys voi edellyttää ymmärrystä siitä, kuinka järjestelmät muuttuvat, kuka niiden muutoksista päättää ja miten järjestelmän muutosten havainnoiminen eroaa niihin tehtävien muutosten tekemisestä.¹⁹⁵

Voidaankin todeta, että läpinäkyvyydelle ei ole vakiintunut yhtä selkeää kiveen hakattua määritelmää, vaan kyseessä on verraten avoin ja abstrakti periaate.¹⁹⁶ Lainsäädännössä on kuitenkin tarkennettu eri kontekstissa avoimuutta ja läpinäkyvyyttä. Erityisesti tietosuoja-asetuksen sisältämä henkilötietojen käsittelyä koskeva läpinäkyvyyden periaate, tietosuoja-asetuksen automaattista päätöksentekoa koskevat suojoitimet ja viranomaisten päätösten perusteluvelvollisuus ovat relevantteja tältä kannalta.¹⁹⁷ Seuraavissa luvuissa analysoidaan edellä mainittujen tietosuoja-asetuksen ja hallinto-oikeudellisten normien avoimuutta ja läpinäkyvyyttä koskevien velvoitteiden toteuttamista tarkemmin.

Algoritmisen päätöksenteon läpinäkyvyyden korostunut merkitys ilmenee myös siinä, että hallinnon automaattista päätöksentekoa koskevaa yleislainsäädäntöä valmisteleva työryhmä on ehdottanut hallituksen esityksen muotoon laatimassaan työryhmämietinnössä hallintolakiin lisättäväksi uuden kohdan, jonka mukaan viranomaisen antamasta kirjallisesta hallintopäätöksestä on käytävä ilmi, jos asia on ratkaistu automaattisesti.¹⁹⁸

Seuraavaksi käsitellään tarkemmin läpinäkyvyyttä henkilötietojen käsittelyn kontekstissa huomioiden, että algoritmiseen päätöksentekoon saattaa sisältyä henkilötietojen käsittelyä, minkä lisäksi tietosuoja-asetus sisältää nimenomaisia automaattista päätöksentekoa koskevia velvoitteita rekisterinpitäjälle.¹⁹⁹

4.2. Henkilötietojen käsittelyn läpinäkyvyys

¹⁹⁴ Ribana – Bohn – Duarte – Garcke 2020, s. 42201.

¹⁹⁵ Ananny – Crawford 2018, s. 984.

¹⁹⁶ Ks. läpinäkyvyyden avoimesta ja abstraktista merkityksestä Hakkarainen ym. 2020, s. 8.

¹⁹⁷ Läpinäkyvyydestä henkilötietojen käsittelyn osalta säädetään tietosuoja-asetuksen 5(1)(a) artiklassa ja automaattisesta päätöksenteosta tietosuoja-asetuksen 22 artiklassa. Päätösten perusteluvelvollisuudesta taas säädetään hallintolain 45 §:ssä.

¹⁹⁸ Oikeusministeriön työryhmämietintö 2022, s. 81.

¹⁹⁹ Ks. esimerkiksi tietosuoja-asetuksen 13–15 artiklat ja 22 artikla.

Hakkaraisen ym. mukaan algoritmisen päätöksenteon avoimuuden voidaan hahmottaa kohdistuvan erityisesti päätöksentekojärjestelmän toimintaperiaatteisiin, mistä johtuen avoimuuden normipohjan muodostavat heidän mukaansa tietosuoja-asetuksen läpinäkyvyyttä, ymmärrettävyyttä ja rekisteröityjen oikeuksia koskevat artiklat sekä automaattisen päätöksenteon hyödyntämisen rajoituksia koskevat artiklat.²⁰⁰

Yksilön kannalta tärkeää on, että hänelle annetaan tarvittavat välineet henkilötietojensa käsittelyn tarkasteluun. Yksilön informoinnilla on olennainen merkitys sille, että yksilö voi nauttia oikeudesta tehdä valintoja, mikä osaltaan taas vahvistaa yksilön autonomiaa. Tästä seuraa, että yksilöillä tulisi olla myös kontrolli heistä tehtäviin profiileihin.²⁰¹

Rekisterinpitäjien kannalta olennaista on ymmärtää, mitä läpinäkyvyysvelvoitteet tarkoittavat käytännössä, ja miten rekisterinpitäjien tulisi toteuttaa läpinäkyvyys.²⁰² Läpinäkyvyyttä ei ole määritelty tietosuoja-asetuksessa, mutta sen merkityksestä saadaan osviittaa tietosuoja-asetuksen johdanto-osan 39 kappaleesta, jonka mukaan²⁰³

''Henkilötietojen käsittelyn olisi oltava laillista ja asianmukaista. Luonnollisille henkilöille olisi oltava läpinäkyvää, miten heitä koskevia henkilötietoja kerätään ja käytetään ja niihin tutustutaan tai niitä käsitellään muulla tavoin sekä selvillä siitä, missä määrin henkilötietoja käsitellään tai on määrä käsitellä.''

Tietosuoja-asetuksessa todetaan lisäksi, että henkilötietoja on käsiteltävä lainmukaisesti, asianmukaisesti ja rekisteröidyn kannalta läpinäkyvällä tavalla.²⁰⁴ Läpinäkyvyys liittyy henkilötietojen käsittelyn asianmukaisuuteen ja osoitusvelvollisuuden periaatteeseen, sillä rekisterinpitäjän tulee kyetä osoittamaan, että henkilötietoja käsitellään rekisteröidyn kannalta läpinäkyvällä tavalla.²⁰⁵ Läpinäkyvyys ilmentää Euroopan unionin perusoikeuskirjan 8 artiklan sisältämää periaatetta, jonka mukaisesti henkilötietojen käsittelyn tulee olla asianmukaista.²⁰⁶ Läpinäkyvyyden toteutuminen on lisäksi olennaista henkilötietojen käsittelyn kohtuullisuuden periaatteen kannalta. Kohtuullisuuden periaatteen mukaisesti tärkeää on se, että rekisteröity saa etukäteen tiedon henkilötietojen käsittelyn

²⁰⁰ Hakkarainen ym. 2020, s. 5. Ks. tietosuoja-asetuksen 13–15 artiklat läpinäkyvyydestä, ymmärrettävyydestä ja rekisteröidyn oikeudesta saada pääsy tietoihin ja 22 artikla automaattisen päätöksenteon rajoitteista.

²⁰¹ Hildebrandt – Gutwirth 2008, s. 122.

²⁰² 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 4.

²⁰³ 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 6.

²⁰⁴ Tietosuoja-asetuksen 5(1)(a) artikla ja tietosuoja-asetuksen johdanto-osan 60 kappale.

²⁰⁵ Tietosuoja-asetuksen 5(2) artikla. Ks. 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 5.

²⁰⁶ 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 5.

laajuudesta ja seurauksista, jotta henkilötietojen käsittelyn tavat eivät yllätä rekisteröityä myöhemmin.²⁰⁷

Algoritmisen päätöksenteon läpinäkyvyydestä ja ymmärrettävyydestä keskusteltaessa on tarpeen erottaa toisistaan henkilötietojen käsittelyn läpinäkyvyys ja algoritmisen päätöksenteon läpinäkyvyys ja ymmärrettävyys. Esimerkiksi tilanne, jossa henkilön henkilötietoja käytetään häntä koskevan päätöksen tekemiseen (esimerkiksi lainahakemuksen myöntäminen tai sosiaalituen myöntäminen) tulee rekisterinpitäjän lähtökohtaisesti informoida rekisteröityä henkilötietojen käsittelystä läpinäkyvällä tavalla tietosuoja-asetuksen 12–14 artiklojen perusteella. Tämän lisäksi rekisterinpitäjän tulee erikseen arvioida, millä tavalla sen tulee selittää automaattista päätöksentekoprosessia rekisteröidylle tietosuoja-asetuksen 13(2)(f), 14(2)(g) ja 15(1)(h) artiklojen perusteella.

Läpinäkyvyyden periaate mahdollistaa sen, että rekisteröidyt voivat saattaa rekisterinpitäjät ja henkilötietojen käsittelijät vastuuseen rikkomuksista ja valvoa omien henkilötietojensa käsittelyä hyödyntämällä heille kuuluvia oikeuksia.²⁰⁸ Läpinäkyvyyden merkitys on tunnustettu myös EU:n tuomioistuimessa, jossa on todettu, että rekisteröityjen informointi on edellytyksenä sille, että rekisteröidyt voivat käyttää tietosuojaoikeuksiaan.²⁰⁹

Toisaalta tietosuoja-asetuksen sisältämät oikeussuojakeinot eivät taas sellaisenaan välttämättä ole riittäviä yksittäisen päätöksen luoman syrjinnän korjaamiseksi. Nimittäin tietosuoja-asetus ei sisällä viranomaisen päätöstä koskevaa valitusoikeutta, vaan tästä säädetään hallinnon kontekstissa sen sijaan oikeudenkäynnistä hallintoasioissa annetun lain 3 luvussa.²¹⁰

Läpinäkyvyys voidaan hahmottaa tietosuoja-asetuksen kontekstissa ennemmin käyttäjakeskeiseksi kuin oikeudelliseksi käsitteeksi. Tämä ilmenee siten, että läpinäkyvyyden käytännön toteuttamisvaatimuksista säädetään tietosuoja-asetuksen artikloissa.²¹¹ Seuraavaksi tarkastellaan tarkemmin tietosuoja-asetuksessa rekisterinpitäjille asetettuja informointivelvoitteita.

4.2.1. Läpinäkyvyys tietosuoja-asetuksessa

²⁰⁷ 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 7.

²⁰⁸ 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 5.

²⁰⁹ Julkisasiamies Cruz Villalónin (C-201/14), kohta 74.

²¹⁰ Tässä tutkielmassa ei tutkita tarkemmin tietosuoja-asetuksen mukaisia oikeussuojakeinoja.

²¹¹ 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 5–6.

Tietosuoja-asetuksen henkilötietojen käsittelyn läpinäkyvyyttä koskevan periaatteen mukaan rekisterinpitäjän tulee informoida rekisteröityä henkilötietojen käsittelyä koskevista seikoista.²¹² Rekisteröidyille toimitettavien tietojen sisällöstä säädetään tietosuoja-asetuksen 13 ja 14 artikloissa, jotka edellyttävät, että rekisterinpitäjä kertoo rekisteröidylle henkilötietojen käsittelystä ja sen tarkoituksista.²¹³ Rekisterinpitäjän on lisäksi annettava rekisteröidylle lisätiedot, jotka ovat tarpeen sen varmistamiseksi, että käsittely on asianmukaista ja läpinäkyvää. Rekisterinpitäjän tulee ilmoittaa myös profiloinnin olemassaolosta ja profiloinnin seurauksista rekisteröidyille.²¹⁴

Tietosuoja-asetuksen mukainen henkilötietojen käsittelyn läpinäkyvyys sisältää kolme osaluettua:

- 1) Henkilötietojen käsittelyä koskevan tiedon toimittaminen rekisteröidyille.
- 2) Rekisterinpitäjien tapa tiedottaa rekisteröidyille tietosuoja-asetuksen sisältämistä oikeuksista.
- 3) Rekisterinpitäjien keinot avustaa rekisteröityjä oikeuksien käyttämisessä.²¹⁵

Seuraavaksi käsitellään tarkemmin tietosuoja-asetuksen 13–14 artikloissa konkretisoituja läpinäkyvyyden periaatteen mukaisia velvoitteita.

Tietosuoja-asetuksen 13–14 artiklat

Tietosuoja-asetuksen 13–14 artikloiden perusteella rekisterinpitäjän tulee toimittaa rekisteröidylle muun muassa tiedot henkilötietojen käsittelyn tarkoituksista, henkilötietojen säilytysajoista ja tietojen mahdollisista vastaanottajista. Tietosuoja-asetuksen 13–14 artikloiden informointivelvoite kohdistuu pääsääntöisesti ajallisesti kohtaan ennen käsittelytoimien (mukaan lukien algoritmisen päätöksenteon) alkamista.²¹⁶

Edellä kuvattujen henkilötietojen käsittelyn niin sanottujen perustietojen lisäksi tietosuoja-asetuksen 13 ja 14 artiklat sisältävät erityisiä automaattista päätöksentekoa ja profilointia koskevia informointivelvoitteita, joiden perusteella rekisterinpitäjien on selitettävä

²¹² 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 4.

²¹³ Tiedonantovelvoitteisiin on eräitä poikkeuksia, jolloin rekisterinpitäjän ei tarvitse toimittaa tietoja rekisteröidylle. Poikkeus tulee kyseeseen, kun henkilötietoja kerätään suoraan rekisteröidyltä ja rekisteröity on jo saanut tiedot (tietosuoja-asetuksen 13(4) artikla). Lisäksi kun tietoja ei ole saatu rekisteröidyltä, informointivelvoitteista voidaan poiketa, jos rekisteröity on jo saanut tiedot, kyseisten tietojen toimittaminen osoittautuu mahdottomaksi tai vaatisi kohtuutonta vaivaa, tietojen hankinnasta tai luovuttamisesta säädetään lainsäädännössä tai tiedot on pidettävä luottamuksellisina, koska niitä koskee lainsäädäntöön perustuva vaitiolovelvollisuus, kuten lakisääteinen salassapitovelvollisuus (tietosuoja-asetuksen 14(5) artikla).

²¹⁴ Tietosuoja-asetuksen johdanto-osan 60 kappale ja tietosuoja-asetuksen 13–14 artiklat.

²¹⁵ 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 4.

²¹⁶ Edwards-Veale 2017, s. 52.

rekisteröidyille selkeästi ja yksinkertaisesti, miten profilointi tai automatisoitu päätöksentekoprosessi toimii: Tietosuoja-asetuksen 13(2)(f) artiklan ja 14(2)(g) artiklan mukaan rekisterinpitäjien tulee kertoa automaattisen päätöksenteon, muun muassa profiloinnin olemassaolo, sekä ainakin näissä tapauksissa antaa merkitykselliset tiedot käsittelyyn liittyvästä logiikasta ja kyseisen käsittelyn merkittävyydestä ja sen mahdollisista seurauksista rekisteröidyille.

WP29 tietosuojatyöryhmän näkemyksen mukaan näiden tietojen toimittamisella rekisterinpitäjät voivat pyrkiä varmistumaan osaltaan, että ne toteuttavat joitakin tietosuoja-asetuksen 22(3) artiklan ja johdanto-osan 71 kappaleen edellyttämiä suojaamistoimenpiteitä. Lisäksi WP29 tietosuojatyöryhmä suosittaa, että vaikka automatisoidut päätökset ja profilointi eivät täyttäisi tietosuoja-asetuksen 22(1) artiklan mukaista määritelmää, olisi tietojen toimittaminen silti hyvän käytännön mukaista.²¹⁷

Merkityksellisillä tiedoilla ”käsittelyyn liittyvästä logiikasta” tarkoitetaan sitä, että rekisterinpitäjän tulisi ottaa käyttöön yksinkertaisia tapoja kertoa rekisteröidyille päätöksen taustalla olevista syistä tai päätöksen tekemiseen hyödynnetyistä perusteista. WP29 tietosuojatyöryhmän mukaan kyseeseen eivät kuitenkaan välttämättä tule monimutkaisen selityksen toimittaminen päätöksenteossa käytettävistä algoritmeista tai koko algoritmin paljastaminen. Tietojen tulee kuitenkin olla riittävän kattavat sisällöltään, jotta rekisteröidyn on mahdollista ymmärtää päätöksen perusteet.²¹⁸

”Merkittävyydellä” ja ”mahdollisilla seurauksilla” viitataan siihen, että rekisterinpitäjän tulee toimittaa tietoja käsittelystä sekä automatisoitujen päätösten mahdollisista vaikutuksista rekisteröityyn. WP29 tietosuojatyöryhmän mukaan tietojen mielekkyys ja ymmärrettävyys edellyttävät, että rekisterinpitäjän tulee antaa todellisia konkreettisia esimerkkejä päätösten mahdollisista vaikutuksista.²¹⁹

WP29 tietosuojatyöryhmä on laatinut ohjeistusta rekisteröidyille annettavan informaation tasosta. Sen mukaan rekisterinpitäjien tulee pohtia selkeitä ja ymmärrettäviä tapoja toimittaa tietoja rekisteröidyille algoritmien tai koneoppimisen toimintatapaa koskevien monimutkaisten matemaattisten selitysten sijasta. Lisäksi visualisointia ja

²¹⁷ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 26–27.

²¹⁸ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 27.

²¹⁹ Esimerkiksi vakuutusyhtiön hyödyntäessä automatisoitua päätöksentekoprosessia, jolla liikennevakuutusmaksut asetetaan henkilöiden ajokäyttäytymisen seurannan tuloksena, voisi vakuutusyhtiö käsittelyn merkittävyyden ja mahdollisten seurausten kuvaamiseksi selittää, että vaarallisen ajotavan seurauksena rekisteröidyille saatetaan määrätä korkeampia vakuutusmaksuja. Vakuutusyhtiö voi tarjota mahdollisuuden sovelluksen kautta vertailla kuvitteellisia kuljettajia erilaisilla ajotavoilla vaarallisuuden osalta vinkkeineen siitä, miten ajotottumuksia voi parantaa vakuutusmaksuja alentaen. Ks. 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 28.

vuorovaikutteisuuden perustuvia tekniikoita algoritmien läpinäkyvyyden edistämiseksi tulisi harkita. Tällaisia suositeltavia tapoja voivat olla esimerkiksi:

- a) Profiloinnissa tai päätöksentekoprosessissa käytetyt tietoryhmät ja tieto siitä, miksi kyseiset tietoryhmät on määritelty merkityksellisiksi.
- b) Tieto siitä, miten automatisoidussa päätöksentekoprosessissa hyödynnetty rekisteröidyn profiili on luotu, sisältäen analyysissa mahdollisesti käytetyt tilastot.
- c) Tieto siitä, miksi laadittu profiili on merkityksellinen automatisoidun päätöksentekoprosessin näkökulmasta.
- d) Tieto siitä, miten profiilia hyödynnetään rekisteröityä koskevan päätöksen tekemisessä.²²⁰

Profilointiin liittyen WP29 tietosuojatyöryhmä on todennut, että etenkin jos tietojenkäsittely sisältää profilointiin perustuvaa päätöksentekoa (riippumatta siitä, onko kyseessä automaattinen päätöksenteko), rekisteröidylle on kerrottava, että käsittelyn tarkoituksena on profilointi, ja että päätöksen tekeminen tapahtuu luodun profiilin perusteella.²²¹

WP29 tietosuojatyöryhmä on lisäksi todennut, että profilointi ja automatisoidut päätökset saattavat aiheuttaa riskejä rekisteröidyn oikeuksille ja vapauksille, mistä johtuen rekisterinpitäjien tulee toteuttaa asianmukaisia suojatoimia.²²²

Tietosuoja-asetuksen 15 artikla

Tietosuoja-asetuksen 15 artiklan mukainen oikeus saada pääsy tietoihin käsittää jo alkaneeseen käsittelytoimeen liittyvien tietojen toimittamisen. Tietosuoja-asetuksen 15 artiklan mukaisten oikeuksien hyödyntäminen mahdollistaa osaltaan rekisteröidyn mahdollisuuden tulla tietoiseksi itseään koskevasta päätöksestä sisältäen myös profilointiin perustuvat päätökset.²²³ Rekisteröidyn tulee kuitenkin aktiivisesti ajaa kyseisen oikeuden toteuttamista suhteessa rekisterinpitäjään, sillä kyseessä on rekisteröidyn oikeus saada pääsy tietoihin, ei rekisterinpitäjän velvollisuus toimittaa tietoja proaktiivisesti.

Tietosuoja-asetuksen 15 artiklan mukaisesti rekisteröidyllä on oikeus saada pääsy hänestä kerättyihin henkilötietoihin, jotta rekisteröity pysyy tietoisena käsittelyn lainmukaisuudesta ja voi tarkistaa sen. Rekisteröidyllä on oikeus tietää ja saada ilmoitus erityisesti henkilötietojen käsittelyn tarkoituksista, ja jos mahdollista, käsittelyajasta, henkilötietojen

²²⁰ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 33–34.

²²¹ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 17–18.

²²² 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 5.

²²³ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 29.

vastaanottajista, käsiteltävien henkilötietojen automaattisen käsittelyn logiikasta sekä kyseisen käsittelyn mahdollisista seurauksista, ainakin jos käsittely perustuu profilointiin.²²⁴

Tietosuoja-asetuksen 15(1)(h) artikla sisältää identtisen muotoilun 13(2)(f) artiklan ja 14(2)(g) artiklan kanssa. 15(1)(h) artiklan mukaan rekisterinpitäjien tulee kertoa automaattisen päätöksenteon, muun muassa profiloinnin olemassaolo, sekä ainakin näissä tapauksissa antaa merkitykselliset tiedot käsittelyyn liittyvästä logiikasta ja kyseisen käsittelyn merkittävyydestä ja sen mahdollisista seurauksista rekisteröidylle.

Rekisteröidyn oikeus saada pääsy tietoihinsa ei kuitenkaan saa vaikuttaa epäedullisella tavalla muiden oikeuksiin ja vapauksiin, joihin kuuluvat esimerkiksi liikesalaisuudet tai henkinen omaisuus ja ohjelmistojen tekijänoikeudet.²²⁵

Edwards ja Veale esittävät, että kyseisen tietosuoja-asetuksen 15 artiklan perusteella rekisterinpitäjä voisi toimittaa taannehtivaa informaatiota tiettyjä yksittäisiä rekisteröityjä koskevista päätöksistä sisältäen yksittäisen päätöksen logiikan tai perusteet, syyt ja yksittäiset olosuhteet.²²⁶ Toisaalta on huomattava, että rekisterinpitäjän olisi tullut jo tässä vaiheessa toimittaa 15 artiklan mukaisia pelkästään automaattiseen käsittelyyn perustuvien päätösten (profilointi mukaan lukien) tietoja osana tietosuoja-asetuksen 13–14 artikloiden mukaista informointivelvoitettaan.²²⁷

Läpinäkyvyyden suhde muihin rekisteröityjen oikeuksiin

Tietosuoja-asetus sisältää myös muita kuin rekisteröidyn informointia koskevia oikeuksia. Eräät tietosuoja-asetuksen oikeudet ovat riippuvaisia tietoihin pääsyä koskevasta oikeudesta kuten tietosuoja-asetuksen 16 artiklan mukainen oikeus tietojen oikaisemiseen.²²⁸

Tietosuoja-asetuksen 16 artiklan mukainen oikeus tietojen oikaisemiseen saattaa tulla kyseeseen esimerkiksi, jos henkilö sijoitetaan ryhmään, joka kertoo jotakin hänen kyvystään suorittaa tietty tehtävä, mutta kyseinen henkilöä koskeva profiili perustuu väärin tietoihin.

²²⁴ Tietosuoja-asetuksen johdanto-osan 63 kappale ja tietosuoja-asetuksen 15(1)(h) artikla, jonka mukaan rekisteröidyllä on oikeus saada seuraavat tiedot ”*automaattisen päätöksenteon, muun muassa 22 artiklan 1 ja 4 kohdassa tarkoitetun profiloinnin olemassaolo, sekä ainakin näissä tapauksissa merkitykselliset tiedot käsittelyyn liittyvästä logiikasta samoin kuin kyseisen käsittelyn merkittävyys ja mahdolliset seuraukset rekisteröidylle.*”

²²⁵ Tietosuoja-asetuksen 15(4) artikla ja tietosuoja-asetuksen johdanto-osan 63 kappale, jossa todetaan, että edellä mainittujen seikkojen huomioon ottaminen ei kuitenkaan saisi johtaa siihen, että rekisteröidylle ei toimiteta lainkaan tietoja. Tässä tutkielmassa ei käsitellä tarkemmin edellä kuvattuja rajoituksia tietosuoja-asetuksen 15 artiklan mukaisten oikeuksien toteuttamiseen.

²²⁶ Edwards-Veale 2017, s. 52.

²²⁷ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 28–29.

²²⁸ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 17.

On varsin luonnollista, että tällaisessa tilanteessa henkilöt haluavat kyseenalaistaa käytettyjen tietojen sekä heitä koskevan ryhmittelyn tai luokittelun paikkansapitävyyden.²²⁹

Oikaisuoikeuden tehokkuus algoritmisen päätöksenteon syrjivyyden osalta riippuu kuitenkin olennaisesti rekisteröidylle annettavien tietojen kattavuudesta ja niiden ymmärrettävyydestä. Voidaan kuvitella tilanne, jossa rekisteröity havaitsee itseään koskevan virheellisen luokittelun (esimerkiksi työtilanteen tai siviilisäädyn suhteen), jolloin rekisteröity voi pyytää tiedon oikaistavaksi. Toisenlainen tilanne on käsillä silloin, jos rekisteröity on luokiteltu tavalla, joka ei varsinaisesti ole virheellinen (esim. rekisteröity on nuori miespuolinen henkilö, mikä on totta), mutta joka algoritmin toteuttaman korrelaatio-suhteisiin perustuvan päättelyn tuloksena johtaa syrjivään lopputulokseen (jos nuoret miespuoliset henkilöt ottavat tekoälyjärjestelmän saatavilla olevan datan perusteella enemmän riskejä, jolloin rekisteröidylle myönnetään jokin etuus huonoimmilla ehdoilla).²³⁰ Tässä tilanteessa rekisteröidyn ei ole mahdollista havaita mitään korjattavaa hänen tiedoissaan (iässään tai sukupuolestaan).

4.2.2. Läpinäkyvyyden toteutustapa tietosuoja-asetuksessa

Tietosuoja-asetuksessa ei aseteta ainoastaan tiedon sisältöä koskevia vaatimuksia, vaan myös toimitettavien tietojen laatu, saatavuus ja ymmärrettävyys ovat tärkeitä seikkoja läpinäkyvyyden varmistamiseksi.²³¹ Tietosuoja-asetuksen 12(1) artiklassa säädetään tarkemmin läpinäkyvyyden toteuttamiseen liittyvien tietojen toimittamistavasta. Rekisterinpitäjän on kyseisen artiklan mukaan osana asianmukaisia toimenpiteitä toimitettava rekisteröidylle 13 ja 14 artiklan mukaiset tiedot ja 15–22 artikloiden ja 34 artiklan mukaiset kaikki käsittelyyn liittyvät tiedot:

- Tiiviillä tavalla esitetyssä, läpinäkyvässä, helposti ymmärrettävässä ja rekisteröidyn saatavilla olevassa muodossa.
- Selkeällä ja yksinkertaisella kielellä.
- Kirjallisesti tai muulla tavoin ja tapauksen mukaan sähköisessä muodossa. Tiedot voidaan antaa suullisesti, jos rekisteröity sitä pyytää ja rekisteröidyn henkilöllisyys on vahvistettu muulla tavoin.²³²

²²⁹ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 19.

²³⁰ On huomioitava, että algoritmin päätelmä voi sinänsä olla totuudenmukainen, mutta sen käyttämistä yleistäen jokaiseen hakijaan voidaan pitää ongelmallisena syrjivyyden kannalta. Ks. esim. Edwards-Veale 2017, s. 30.

²³¹ 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 6.

²³² Tietosuoja-asetuksen 12(1) artikla.

Täyttääkseen vaatimuksen siitä, että rekisteröidyille suunnatun tiedon ja viestinnän on oltava ”tiivisti esitetyssä ja läpinäkyvässä” muodossa, tulee rekisterinpitäjien esittää tiedot tehokkaasti ja ytimekkäästi, välttäen rekisteröidyn kuormittamista liialla tiedolla. Helposti ymmärrettävän muodon vaatimus taas merkitsee sitä, että kohdeyleisön keskimääräisen edustajan tulisi kyetä ymmärtämään, mitä tiedoissa tarkoitetaan. Ymmärrettävyys ja selkeän ja yksinkertaisen kielen vaatimukset liittyvät läheisesti toisiinsa.²³³

Tietosuojasetuksen yhteydessä tunnistettu sisäinen jännite ilmenee siinä, että tietosuojasetuksessa toisaalta edellytetään antamaan rekisteröidyille kattavasti asetuksessa edellytetyt tiedot, mutta tiedot on kuitenkin toimitettava rekisteröidyille tiiviisti esitetyssä, läpinäkyvässä, helposti ymmärrettävissä olevassa ja saatavilla olevassa muodossa. Rekisterinpitäjän vastuulle jääkin tapauskohtaisesti arvioida toteuttamansa henkilötietojen käsittelyn luonnetta, olosuhteita, laajuutta ja asiayhteyttä. Tämän arvion pohjalta rekisterinpitäjän tulee päättää, miten yksityiskohtaisesti ja millä tavoilla tiedot toimitetaan rekisteröidyille.²³⁴

Selkeän ja yksinkertaisen kielen vaatimuksen mukaan rekisterinpitäjän tulee antaa tiedot mahdollisimman yksinkertaisessa muodossa ja välttää monimutkaisten lauserakenteiden ja monimutkaisen kielen käyttöä. Lisäksi tietojen tulisi olla konkreettisia ja lopullisia, eikä niitä tulisi esittää abstraktien, epäselvien tai monitulkintaisten keinojen avulla.²³⁵

Läpinäkyvyyden periaatteen mukaisesti rekisteröidyn tulee saada etukäteen tieto käsittelyn laajuudesta ja seurauksista, eivätkä henkilötietojen käyttötavat saisi tulla rekisteröidyille myöhemmin yllätyksenä. Edellä mainittu läpinäkyvyys on merkittävää myös tietosuojasetuksen 5 artiklan mukaisen kohtuullisuuden periaatteen toteutumisen kannalta. WP29 tietosuojatyöryhmän mukaan etenkin tietojen käsittelyn ollessa monimutkaista, teknistä tai odottamatonta, tulee rekisterinpitäjän tietosuojasetuksen 13 ja 14 artiklan mukaisten tietojen lisäksi selittää erikseen ja selkeää kieltä käyttäen henkilötietojen käsittelyn tärkeimmät seuraukset, eli henkilötietojen käsittelyn vaikutukset rekisteröityyn.²³⁶

²³³ 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 7.

²³⁴ 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 18.

²³⁵ 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 8–9.

²³⁶ 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 7.

On tärkeää huomata, ettei monimutkaisuus kuitenkaan ole peruste jättää toimittamatta tietoja rekisteröidylle.²³⁷ Lisäksi rekisterinpitäjän tulisi pidättäytyä käyttämästä informoinnissa liian oikeudellista, teknistä tai erikoisalojen kieltä tai sanastoa.²³⁸

Rekisterinpitäjien tulee siis kyetä toimittamaan riittävän kattavat tiedot automaattisesta päätöksenteosta samalla varmistuen, että rekisteröidyn on mahdollista tietojen perusteella muodostaa riittävä ymmärrys algoritmista päätöksenteosta ja siinä tapahtuvasta henkilötietojen käsittelystä.

4.3. Tietosuoja-asetuksen mukainen automaattisen päätöksenteon selitysoikeus

Edellä kuvattiin läpinäkyvyyden merkitystä ja toteuttamistapoja, jotta yksilöillä olisi mahdollisuus valvoa heitä koskevaa algoritmista päätöksentekoa. On kuitenkin kysyttävä, seuraako jonkin tiedon vastaanottamisesta automaattisesti tietosisällön ymmärtäminen merkityksellisellä tavalla. Hakkarainen ym. ovatkin sitä mieltä, että läpinäkyvydessä on kyse vain illuusiosta, jolloin asian läpinäkyvyydestä ei automaattisesti seuraa tarkasteltavan asian ymmärtämistä.²³⁹

Tässä tutkielmassa läpinäkyvydellä viitataan aineiston tai tiedon toimittamiseen tai saataville saattamiseen rekisteröidylle, kun taas ymmärrettävyydellä viitataan siihen, että rekisteröity voi todella ymmärtää hänelle toimitetun tiedon. Läpinäkyvyyden nojalla toimitettujen tietojen ymmärrettävyys taas saattaa edellyttää tiedon sisällön ja merkityksen selittämistä auki tarkemmin rekisteröidylle.²⁴⁰ Tässä tutkielmassa käytetään kuitenkin laajemmassakin merkityksessä läpinäkyvyys-termiä, silloin kun lähdeaineistossa on käytetty kyseistä termiä.

Burrellin mukaan algoritmisen päätöksenteon ymmärrettävyyden esteenä eivät välttämättä toimi pelkkä tietoaineiston suuri määrä tai koodi, sillä ainakin teoriassa suuretkin tietoaineistot voidaan esittää ymmärrettävällä tavalla ja koodia on mahdollista kirjoittaa selkeästi. Haasteen ymmärrettävyydelle aiheuttaa usein tietoaineistojen ja koodin yhteen toimiminen algoritmin toimintamallissa.²⁴¹ Tiedolliset ja voimavaroihin liittyvät epäsymmetriat rekisteröidyn ja rekisterinpitäjän välillä eivät Janssenin mukaan johdu ainoastaan koneoppimisen läpinäkymättömyydestä, vaan niitä ruokkii myös se, että ihmisillä ei usein algoritmiseen päätöksentekojärjestelmään liittyen ole ymmärrystä siitä, miten

²³⁷ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 27.

²³⁸ 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 10.

²³⁹ Hakkarainen ym. 2020, s. 8.

²⁴⁰ Ks. selittämisen merkityksestä O'Hara 2020, s. 2–4.

²⁴¹ Burrell 2016, s. 5.

organisaatiot käsittelevät heidän tietojansa. Heillä ei myöskään usein ole ymmärrystä kyseessä olevasta teknisestä arkkitehtuurista, tietojen jakamisesta tai järjestelmiin liittyvistä liiketoimintamalleista.²⁴²

Algoritmisen päätöksenteon osalta tämä tarkoittaisi sitä, että yksilö tietää miksi päätös tehtiin, mutta ei ymmärrä päätöksen syitä.²⁴³ Voidaankin todeta, että algoritmiin liittyvän datan toimittamisella rekisteröidylle on vähäinen merkitys rekisteröidyn oikeuksien turvaamiselle, mikäli rekisteröidyn ei ole mahdollista ymmärtää algoritmia ja sitä kautta haastaa epäreiluja algoritmeja.²⁴⁴

Tietämisen ja ymmärtämisen ero saattaa joissakin tilanteissa olla häilyvä, mutta kuten O'Hara esittää, jonkin asian selittäminen edellyttää esityksen tai päättelymallin toimittamisen lisäksi tukea tai taustoitusta riippuen rekisteröidyn ymmärryksen tasosta.²⁴⁵

Käytännössä tietosuoja-asetuksen periaatteet jättävät runsaasti tilaa tulkinnalle.²⁴⁶ Voidaankin pohtia, miten hyvin tietosuoja-asetus turvaa ylipäänsä rekisteröityjen mahdollisuuden saada tietoa algoritmisen päätöksenteon mahdollisen syrjivyyden tunnistamiseen.

Kuten edellä on kuvattu, tietosuoja-asetuksen 22 artiklassa kuvataan automaattista päätöksentekoa koskevia suojatoimia, mutta siinä ei tarkemmin yksilöidä, mitä suojatoimet edellyttävät tapauksissa, joissa päätöksenteosta säädetään laissa.²⁴⁷ Toisaalta WP29 tietosuojatyöryhmä on pyrkinyt ohjeistamaan suojatoimista ja informoinnin toteuttamisesta edellä kuvatulla tavalla, mutta kyseinen WP29 tietosuojatyöryhmän materiaali ei sisällä erityisen yksityiskohtaista teknisen tason ohjeistusta rekisterinpitäjille.

Akateemisessa keskustelussa on pohdittukin sitä, muodostavatko tietosuoja-asetuksen läpinäkyvyyttä koskevat velvoitteet yhdessä tietosuoja-asetuksen automaattista päätöksentekoa koskevien velvoitteiden kanssa selitysoikeuden (niin sanottu *'right to explanation'*) algoritmista päätöksentekoa käytettäessä. Myös rekisteröidyn oikeutta saada pääsy itseään koskeviin henkilötietoihin on arvioitu selitysoikeuden kannalta.²⁴⁸

²⁴² Janssen 2020, s. 82.

²⁴³ O'Hara 2020, s. 3.

²⁴⁴ Pasquale 2015, s. 22.

²⁴⁵ O'Hara 2020, s. 3.

²⁴⁶ Janssen 2020, s. 79.

²⁴⁷ Oikeusministeriön arviomuistio 2020, s. 57.

²⁴⁸ Koivisto 2020, s. 16.

Algoritmisen päätöksenteon selitysoikeudesta ei kuitenkaan suoraan säädetä tietosuojasetuksessa ja sen olemassaolon puolesta ja vastaan on esitetty argumentteja akateemisessa keskustelussa.²⁴⁹ Argumentit vaikuttavat painottuvan tietosuojasetuksen sisällön tulkintaan. Tässä tutkielmassa ei arvioida selitysoikeuden olemassaolon perusteita tarkemmin, vaan keskitytään tilanteisiin, joissa rekisterinpitäjä on katsonut tarpeelliseksi selittää algoritmista päätöksentekoa rekisteröidylle muun toimittamansa tietosuojasetuksen mukaisen informaation ohella.²⁵⁰

Huomioiden selitysoikeuden olemassaolon osalta vallitseva tulkinnanvaraisuus, on ymmärrettävää, ettei selitykselle ole osoitettavissa tarkkarajaista määritelmää algoritmisen päätöksenteon ymmärrettävyyden kontekstissa. Selityksille akateemisessa keskustelussa määritelty laajuus ja toisaalta syy-seuraussuhteen aste vaihteleeekin. Roscher ym. ovat ehdottaneet tekoälyn selityksen sisältävän kuvauksen siitä, mitä päätöksenteossa tapahtui. Tämän jälkeen tulisi kysyä, miten jokin asia tapahtui. Lopuksi tulisi kysyä, miksi jokin asia tapahtui. Toisin sanoen koneoppimisjärjestelmien osalta järjestelmän käyttäjän tavoitteet ovat erittäin olennaisia. Erityisesti tieteellisellä tasolla ymmärrettävät selitykset edellyttävät sitä, että dataa, menetelmien läpinäkyvyyttä ja niiden tulkintaa hyödynnetään selittämään tietty syötteen tulos tai data. Heidän mukaansa toisaalta riittävä selitys on mahdollisesti saavutettavissa järjestelmien valvonnan tarkoituksiin esimerkkien tai paikallisten ominaisuuksien arvioinnin kautta, jotta päätöksenteon kannalta tiettyjä formaaleja näkökulmia on mahdollista tarkastella.²⁵¹

Akateemisesta keskustelusta on havaittavissa myös muita tapoja lähestyä algoritmisen päätöksenteon ymmärrettävyyttä laajemmin, kuin pelkän läpinäkyvyyden toteuttamisen kautta. Malgieri ja Comandé ovat ehdottaneet tietosuojasetuksesta johdettavissa olevaa lukukelpoisuuden (legibility) konseptia, jolla on heidän mukaansa mahdollista ratkaista informointioikeuden ja selitysoikeuden välillä vallitseva ristiriita. Lukukelpoisuudella he viittaavat yksilön kyvykkyyteen itsenäisesti ymmärtää algoritmisen päätöksenteon logiikka, merkitys ja mahdolliset seuraukset.²⁵²

Lukukelpoisuus poikkeaa Malgierin ja Comandén mukaan pelkästä datan tai analytiikan luettavuudesta (readability), koska se sisältää enemmän yksityiskohtia päätöksenteon

²⁴⁹ Brkan 2018, s. 14.

²⁵⁰ Selitysoikeuden olemassaolon puolesta ks. esimerkiksi Brkan 2018. Selitysoikeuden olemassaoloa vastaan koskevan argumentoinnin osalta ks. esimerkiksi Wachter ym. 2017.

²⁵¹ Roscher ym. 2020, s. 42203.

²⁵² Malgieri – Comandé 2017, s. 250.

tarkoituksista, lopullisuudesta, kaupallisesta merkityksestä ja mahdollisista seurauksista. Lukukelpoisuus eroaa myös selitettävyydestä/informoinnista, koska se on proaktiivisempaa ja on räätälöity yksilön henkilökohtaisen ymmärryksen tarkoituksiin ja logiikan sekä seurausten konkreettiseen ymmärrettävyyteen.²⁵³

Lukukelpoisuus on mahdollista Malgierin ja Comandén mukaan saavuttaa kertomalla algoritmisen päätöksenteon yhteydessä merkitykselliset tiedot käsittelyyn liittyvästä logiikasta samoin kuin päätöksenteon merkittävyys ja mahdolliset seuraukset rekisteröidylle. He esittävät, että merkitykselliset-termin tulisi tulkita sisältävän sekä merkittävät tiedot (läpinäkyvyyden edistäminen), että hyödylliset tiedot (ymmärrettävyyden edistäminen).²⁵⁴

4.4. Päätösten perusteluvelvollisuus

Edellä kuvattujen tietosuoja-asetuksen sisältämien henkilötietojen käsittelyn läpinäkyvyyttä edistämään pyrkivien velvoitteiden lisäksi viranomaistoimintaan kohdistuu erityisiä päätösten perustelua koskevia velvoitteita. Perusteluvelvollisuus koskee viranomaisia näiden päätöksenteon yhteydessä. Oikeus saada perusteltu päätös muodostaa osan hyvän hallinnon takeita.²⁵⁵ Kyseinen oikeus ilmaistaan perustuslain 21 §:n 2 momentista:

”Käsittelyn julkisuus sekä oikeus tulla kuulluksi, saada perusteltu päätös ja hakea muutosta samoin kuin muut oikeudenmukaisen oikeudenkäynnin ja hyvän hallinnon takeet turvataan lailla.”

Perustuslain takaamaa oikeutta saada perusteltu päätös konkretisoi hallintolaki, jonka mukaan viranomaisen on perusteltava päätöksensä. Päätöksen perusteluista tulee käydä ilmi ratkaisuun vaikuttaneet seikat ja selvitykset sekä sovelletut säännökset.²⁵⁶

Poikkeaminen perusteluvelvollisuudesta on mahdollista hallintolain 45 §:n sisältämissä tilanteissa. Poikkeaminen on mahdollista tilanteissa, joissa perusteluiden toimittamista ei pidetä asianosaisen oikeusturvan näkökulmasta ehdottoman välttämättömänä ja perusteluiden antamatta jättämisen puolesta puhuvat painavat perusteet.²⁵⁷

²⁵³ Malgieri – Comandé 2017, s. 250.

²⁵⁴ Malgieri – Comandé 2017, s. 265.

²⁵⁵ Mäenpää hallinto-oikeus, sähköinen teos, III hallintomenettely ja hyvän hallinnon perusteet, 8. päätösvallan käyttö ja hallintopäätös, hallintopäätöksen perustelut, perustelujen tehtävät.

²⁵⁶ Hallintolaki 45 §.

²⁵⁷ Mäenpää hallinto-oikeus, sähköinen teos, III hallintomenettely ja hyvän hallinnon perusteet, 8. päätösvallan käyttö ja hallintopäätös, hallintopäätöksen perustelut, poikkeukset perusteluvelvollisuudesta.

Päätösten perusteluvelvoite on relevantti myös viranomaisen hyödyntäessä algoritmista päätöksentekoa. Oikeusministeriön vuonna 2020 julkaisemassa arviomuistiossa hallinnon automaattiseen päätöksentekoon liittyvistä yleislainsäädännön tarpeista analysoidaan perusteluvelvollisuuden toteutumista automaattisen päätöksenteon yhteydessä. Muistiossa todetaan, että automaattista päätöksentekoa koskevassa mahdollisessa yleislainsäädännössä tulisi velvoittaa viranomaista esittämään kuvaus automaattisessa päätöksenteossa sovellettavista päätöksentekosäännöistä ja siitä, miten asiat valikoituvat automaattisesti ratkaistavaksi. Lisäksi asianosaiselle tulisi ilmoittaa päätöksen yhteydessä asian ratkaisemisesta automaattisesti. Arviomuistion mukaan velvoite täydentäisi hallintolain 45 §:ssä säädettyä hallintopäätöksen perustelua koskevaa velvoitetta.²⁵⁸

Päätösten perusteluvelvollisuudelle on mahdollista tunnistaa useita eri tarkoituksia. Ensinnäkin perusteluvelvollisuus on keskeisessä roolissa asianosaisen oikeusturvan kannalta. Asianosaisen on tarpeen saada tietää häntä koskevan ratkaisun tekemiseen vaikuttaneet seikat. Riittävän tarkat perustelut antavat asianosaiselle myös edellytykset harkita muutoksenhakua.²⁵⁹ Lisäksi perusteluiden avulla on mahdollista varmistaa, että viranomainen on käyttänyt harkintavaltaansa lainsäädännössä asetettujen rajoitusten ja hallinnon oikeusperiaatteiden kannalta oikealla tavalla, millä voi olla viranomaisen toiminnan asianmukaisuutta ja puolueettomuutta kohtaan tunnettua luottamusta vahvistava vaikutus.²⁶⁰

Päätösten perusteluvelvollisuuden lisäksi viranomaisten tulee osana toimintaansa huomioida muitakin hallintolain sisältämiä periaatteita, joista olennaisia algoritmisen päätöksenteon läpinäkyvyyden ja ymmärrettävyyden kontekstissa ovat palveluperiaate ja palvelun asianmukaisuus (7 §), neuvontaa koskevat vaatimukset (8 §) ja hyvän kielenkäytön vaatimukset (9§).

Viranomaisille voikin edellä mainituista periaatteista olla johdettavissa velvollisuus kuvata ja selventää automaattista menettelyä sekä menettelyn sääntöjä ja toimintaperiaatteita rekisteröidyille.²⁶¹

4.4.1. Perusteluvelvollisuuden toteuttaminen algoritmisen päätöksenteon osalta

²⁵⁸ Oikeusministeriön arviomuistio 2020, s. 46.

²⁵⁹ HE 72/2002 vp, Hallituksen esitys eduskunnalle hallintolaiksi ja laiksi hallintolainkäyttölain muuttamisesta, s. 100.

²⁶⁰ Mäenpää hallinto-oikeus, sähköinen teos, III hallintomenettely ja hyvän hallinnon perusteet, 8. päätösvallan käyttö ja hallintopäätös, hallintopäätöksen perustelut, perustelujen tehtävät.

²⁶¹ Oikeusministeriön arviomuistio 2020, s. 34.

Viranomaisten tulee kyetä toteuttamaan päätösten perusteluvollisuus myös algoritmisen päätöksenteon yhteydessä.²⁶² Olennainen kysymys on, millaisia perusteluiden tulisi sisältöltään ja laadultaan olla algoritmisen päätöksenteon yhteydessä, jotta ne voidaan katsoa riittäviksi.²⁶³ Toisin sanoen tulee pohtia, miten pitkälle ja teknisesti viranomaisen on velvoitettu kuvaamaan automaattista päätöksentekoa osana perusteluvollisuutta.²⁶⁴

Mäenpään mukaan päätösten perusteluissa tulisi ilmetä selvitettyiksi katsotut seikat ja niille annettu oikeudellinen merkitys, koska kyseisillä tiedoilla on päätöksen kohteen oikeussuojan kannalta olennainen merkitys.²⁶⁵ Lisäksi hänen mukaansa perusteluissa tulisi kuvata päätöksen perustan muodostavat selvitykset ja tosiasioihin liittyvä informaatio.²⁶⁶

Soininen taas on kuvailut päätösten perusteluiden sisältöä seuraavaksi: *“voimassa oleva oikeus velvoittaa lainsoveltajia säännösviittauksen esittämiseen ja säännöksen sanamuodon selostamiseen.”* Soinisen mukaan kaikki edellä mainittua pidemmälle ulottuva perusteluvollisuus määräytyy käytettävissä olevan päätöksentekijän harkintavallan laajuuden perusteella siten, että laaja harkintavalta edellyttää myös laajempaa avoimuusvaatimusta ja syvemmälle ulottuvia perusteluita.²⁶⁷ Hakkaraisen ym. mukaan perusteluiden laajuuteen vaikuttavat myös päätöksentekijän toimivalta ja yksittäisen tapauksen erityispiirteet.²⁶⁸

Voutilainen on esittänyt, että etuihin tai oikeuksien myöntämiseen liittyvissä kielteisissä päätöksissä sekä päätöksissä, joissa asetetaan velvoitteita, tulisi olla selkeät perustelut, jotka ovat juuri tiettyyn ratkaisuun vaikuttaneet. Lisäksi ratkaistaessa kilpailutilanne tai asetettaessa päätökselle ehtoja, taikka päätöksen ollessa osittain myönteinen, tulisi viranomaisen esittää yksilöityjä perusteluita päätökselle. Näissä tapauksissa korostuu yksilöityjen perusteluiden asema yksilön oikeusturvan osana.²⁶⁹

Arvioitaessa sitä, voidaanko perusteluvollisuudella varmistaa rekisteröityjen riittävä ymmärrys tunnistamaan päätöksenteon mahdollista syrjivyyttä, vaikuttaa edellä mainittujen

²⁶² Hallinnon algoritmisen päätöksenteon perusteluvollisuutta on arvioitu ministeriötasolla. Oikeusministeriön vuonna 2020 julkaisemassa arviomuistiossa hallinnon automaattiseen päätöksentekoon liittyvistä yleislainsäädännön tarpeista nostettiin esille perusteluvollisuus osana hallinnon algoritmista päätöksentekoa. Arviomuistiossa ei kuitenkaan otettu kantaa päätösten perusteluiden tarkkoihin teknisiin sisältövaatimuksiin suhteessa automaattiseen päätöksentekoon (ks. oikeusministeriön arviomuistio 2020, s. 40–41).

²⁶³ Hakkarainen ym. 2020, s. 37.

²⁶⁴ Yleisemmin avoimuusvaatimuksesta ks. Hakkarainen ym. 2020, s. 4.

²⁶⁵ Mäenpää hallinto-oikeus, sähköinen teos, III hallintomenettely ja hyvän hallinnon perusteet, 8. päätösvallan käyttö ja hallintopäätös, hallintopäätöksen perustelut, perustelujen sisältö, selkeät, yksilöidyt ja ymmärrettävät perustelut.

²⁶⁶ Mäenpää hallinto-oikeus, sähköinen teos, III hallintomenettely ja hyvän hallinnon perusteet, 8. päätösvallan käyttö ja hallintopäätös, hallintopäätöksen perustelut, perustelujen sisältö, ratkaisuun vaikuttaneet seikat ja selvitykset.

²⁶⁷ Soininen 2017, s. 1147.

²⁶⁸ Hakkarainen ym. 2020, s. 37–38.

²⁶⁹ Voutilainen 2008, s. 27.

seikkojen johdosta siltä, ettei perusteluvollisuus riittäisi ainakaan sellaisenaan takaamaan riittävää ymmärrystä yksilöille päätöksenteon syrjivyyden tunnistamiseen. Kuten edellä on todettu, päätöksen perusteluissa on hallintolain 45 §:n mukaan ilmoitettava, mitkä seikat ja selvitykset ovat vaikuttaneet ratkaisuun sekä mainittava sovelletut säännökset. Tämä viittaisi nähdäkseni siihen, että rekisteröidylle olisi kerrottava sovelletut säädökset ja päätökseen vaikuttaneet seikat, mutta ei kuitenkaan teknistä päätöksentekoprosessia tai sen taustalla vaikuttavia tekijöitä, jotka saattavat perustua esimerkiksi datassa piileviin syrjivyyttä aiheuttaviin vinoumiin (toisin sanoen, miten ratkaisu on teknisesti tehty).

Edellä kuvattua johtopäätöstä tukee hallinnon automaattista päätöksentekoa koskevaa yleislainsäädäntöä valmistelevan työryhmän käyttöalaa ja läpinäkyvyyttä koskevissa säännösluonnoksissa kuvattu läpinäkyvyyden toteuttamistapa. Säännösluonnosten mukaan viranomaisen tulee julkaista hallintoasioiden automaattisesta käsittelystä kuvaus, joka sisältää keskeiset tiedot asioiden automaattiseen käsittelyyn valikoitumisen edellytyksistä ja niiden käsittelyä koskevista oikeudellisista perusteista. Asianosaisen tulisi kyetä kuvauksen avulla ymmärtämään automaattisen käsittelyn keskeiset lähtökohdat sekä sen, miltä osin eri asiaryhmiä on automatisoitu ja millä tavoin asioita valitaan automaattisen käsittelyn piiriin. Säännös ei kuitenkaan edellyttäisi tarkoista algoritmeista laadittavaa yksityiskohtaista tai teknistä kuvausta, vaan automaation ymmärtämisen kannalta keskeisten seikkojen kuvausta.²⁷⁰

Perusteluvollisuuden suhdetta algoritmisen päätöksenteon ymmärrettävyyteen on tarkasteltu myös lainsäädäntövaiheessa, kun Migri-hankkeen yhteydessä algoritmin julkaisemista ymmärrettävässä muodossa ei pidetty varsinaisen perusteluvollisuuden vaatimukset täyttävänä, vaan ennemminkin sitä täydentävänä. Hakkaraisen ym. mukaan yhteys algoritmisen päätöksenteon ymmärrettävyyksivaatimuksen ja perusteluvollisuuden välillä on kuitenkin melko selkeästi havaittavissa, sillä algoritmisten päätösten perustelut ja

²⁷⁰ Hallinnon automaattista päätöksentekoa koskevaa yleislainsäädäntöä valmistelevan työryhmän muistio 2021, s. 7. Tässä yhteydessä tulee huomata, että säännösluonnosten mukaan ehdotetun käyttöalasäännöksen perusteella hallintoasia on sallittua ratkaista automaattisesti vain niissä tapauksissa, joissa ratkaisu on mahdollista johtaa yksiselitteisesti ratkaisuun vaikuttavista seikoista ja selvityksistä sekä oikeussäännöistä. Tällöin automaattisessa päätöksenteossa ei säännösluonnosten perusteella vaikuttaisi olevan sellaisia piiloon jääviä perusteita, joita ei kuvattaisi jo päätöksen perusteluissa, eikä päätöksentekosääntöjen kuvaaminen toisi esimerkiksi oikeusturvan tai päätöksen ymmärtämisen kannalta merkityksellistä lisätietoa. Päätöksentekosääntöissä tulisi kuvata jokseenkin samansisältöiset asiat kuin päätösten perusteluissa. Säännösluonnoksissa ei siten nähty tarpeelliseksi säätää hallinto-oikeudellista velvoitetta koskien hyödynnettyjen päätöksentekosääntöjen yksityiskohtaista kuvaamista hallintopäätöksessä. Ks. Hallinnon automaattista päätöksentekoa koskevaa yleislainsäädäntöä valmistelevan työryhmän muistio 2021, s. 7.

ymmärrettävässä muodossa esitetty algoritmi pyrkivät molemmat selittämään järjestelmän päätöksentekoa maallikolle.²⁷¹

Päätösten perusteluvelvollisuus on tärkeä osa hyvän hallinnon edellytysten toteutumista myös algoritmisen päätöksenteon yhteydessä. Kuitenkin voidaan argumentoida, että pelkästään päätösten perusteluvelvollisuus ei riitä takaamaan riittävää ymmärrystä päätöksenteosta rekisteröidylle, jotta hänen olisi mahdollisuus tunnistaa päätöksenteossa mahdollisesti vallitseva syrjivyyttä.

Tämä johtuu siitä, että päätösten perusteluvelvollisuus keskittyy päätöstä koskevien seikkojen oikeudellisen merkityksen ja selvitettyjen asioiden kuvaamiseen.²⁷² Päätösten perusteluvelvollisuudessa ei varsinaisesti olekaan kyse aktiivisesta päätöksenteon teknisten yksityiskohtien selittämisestä, vaan kyseeseen tulee enemminkin läpinäkyvyyteen vertautuva tiettyjen tietojen toimittaminen rekisteröidylle.

4.5. Algoritmista päätöksentekoa koskevien tietojen toimittaminen suhteessa päätöksenteon elinkaareen

Tässä alaluvussa käsitellään algoritmisen päätöksenteon läpinäkyvyyden toteuttamista algoritmisen päätöksenteon elinkaaren kannalta tarkasteltuna.

Algoritmisen päätöksentekojärjestelmän syrjivyyttä on mahdollista lähestyä purkamalla järjestelmä ikään kuin osiin: ensiksi voidaan tarkastella järjestelmän kehitysvaihetta (miten järjestelmä on kehitetty). Tämän jälkeen voidaan tarkastella sitä, miten järjestelmä käytännössä toimii (mikä järjestelmän logiikka on). Lopuksi tarkastelu voidaan kohdistaa yksittäiseen päätökseen (mitä järjestelmä on päättänyt yksittäistä henkilöä koskien).²⁷³

Akateemisessa keskustelussa on pyritty tunnistamaan rekisteröidylle annettavien tietojen sisältö algoritmisen päätöksenteon selitysoikeuden kannalta tarkasteltuna. Brkanin mukaan merkityksellinen tieto logiikasta käsittäisi ideaalitalanteissa: 1) tiedot syötteenä olevasta datasta 2) tiedot tekijöistä, jotka vaikuttivat päätökseen 3) tiedot päätökseen vaikuttaneiden tekijöiden keskinäisestä merkityssuhteesta ja 4) selityksen siitä, miksi tietty päätös tehtiin. Brkan tunnistaa haasteita 2) ja 3) vaatimusten osalta, minkä tuloksena selitysoikeus

²⁷¹ Hakkarainen ym. 2020, s. 39.

²⁷² Ks. perusteluiden sisällöstä Mäenpää hallinto-oikeus, sähköinen teos, III hallintomenettely ja hyvän hallinnon perusteet, 8. päätösvallan käyttö ja hallintopäätös, hallintopäätöksen perustelut, perustelujen sisältö, selkeät, yksilöidyt ja ymmärrettävät perustelut.

²⁷³ Ks. selitysten kohteesta Hakkarainen ym. 2020, s. 39.

rajautuisi todennäköisesti kattamaan kirjalliset tiedot, jotka selittävät päätökseen vaikuttaneet merkittävät syyt.²⁷⁴

Algoritmista päätöksentekoa koskevien tietojen antamisen ajankohta on merkittävä rekisteröidyn kannalta tarkasteltuna, sillä algoritmiseen päätöksentekoon saattaa sisältyä syrjivyyttä eri vaiheessa päätöksenteon elinkaarta.²⁷⁵ Hakkaraisen ym. mukaan avoimuuden eri toteutustapoja punnittaessa tulee arvioida avoimuuden toteutustavan vaikutusta järjestelmien kontrollin tehokkuuteen ja millaista kontrollia eri avoimuuden tavoilla on mahdollista toteuttaa.²⁷⁶

Toisaalta keskusteltaessa läpinäkyvyydestä algoritmisten päätöksentekojärjestelmien syrjivyyden tunnistamista edistävänä tekijänä, on tärkeää pohtia sitä, mitä ylipäänsä läpinäkyvyyden avulla pyritään avaamaan rekisteröidylle: tuleeko kyseeseen järjestelmän toiminnan, henkilötietojen käsittelytoimien tai yksittäistä päätöstä koskevien parametrien avaaminen?²⁷⁷

On myös huomattava, että tietojen toimittaminen ei sellaisenaan merkitse sitä, että rekisterinpitäjä laatisi selityksen siitä, sisältyykö järjestelmään syrjivyyttä vai ei. Selitys edellyttäisi tietyn päätöksentekoon liittyvän tiedon toimittamisen lisäksi sen merkityksen selittämisen. Ideaalitulanteessa rekisteröity voisikin tunnistaa syrjivyyden selityksen perusteella.

Tietosuoja-asetuksen ja hallintolain sisältämien rekisteröidyn informointia koskevien toimien tehokkuus riippuu rekisterinpitäjän kyvykkyydestä ja halukkuudesta toteuttaa kyseiset velvoitteensa. Mikäli rekisterinpitäjä ei toteuta riittävällä tavalla kyseisiä velvoitteita, jää rekisteröidyn tiedonsaanti lopulta riippuvaiseksi tämän omasta aktiivisuudesta, tietotasosta ja valppaudesta.²⁷⁸

Seuraavaksi tarkastellaan edellä kuvattujen tietosuoja-asetuksen ja hallintolain tietojen toimittamista koskevien velvoitteiden sisältöä suhteessa päätöksenteon elinkaareen.

4.5.1. Ennen päätöksen tekemistä rekisteröidylle annettavat tiedot

Ennen päätöksen tekemistä annettavilla tiedoilla rekisteröity voi ideaalitulanteessa tulla tietoiseksi tulevaisuudessa tapahtuvan päätöksenteon potentiaalisista syrjivyyden

²⁷⁴ Brkan 2018, s. 15.

²⁷⁵ Janssen 2020, s. 82.

²⁷⁶ Hakkarainen ym. 2020, s. 4.

²⁷⁷ Ks. avoimuuden kohteesta Hakkarainen ym. 2020, s. 4.

²⁷⁸ Hildebrandt 2012, s. 51.

aiheuttamista kielteisistä seurauksista. Hän saattaa esimerkiksi kieltäytyä päätöksenteosta tai saattaa asian valvontaviranomaisen käsittelyyn. Tämä on kuitenkin riippuvaista siitä, että rekisteröity saa tarvittavat tiedot ja toisaalta siitä, että hän kykenee tietojen pohjalta tunnistamaan mahdollisen syrjivyyden. Rekisterinpitäjän näkökulmasta taas olennaista on tieto siitä, mitä tietoja rekisterinpitäjän tulee toimittaa noudattaakseen sille asetettuja informointivelvoitteita.²⁷⁹

Tietosuoja-asetuksessa velvoitetaan rekisterinpitäjä toimittamaan tiettyjä algoritmista päätöksentekoa koskevia tietoja rekisteröidylle. Tietosuoja-asetuksen 13(2)(f) artiklan ja 14(2)(g) artiklan mukaan rekisterinpitäjien tulee kertoa automaattisen päätöksenteon, muun muassa profiloinnin olemassaolo, sekä ainakin näissä tapauksissa antaa merkitykselliset tiedot käsittelyyn liittyvästä logiikasta ja kyseisen käsittelyn merkittävyydestä ja sen mahdollisista seurauksista rekisteröidylle.

Rekisterinpitäjien tulisi yllä kuvattujen velvoitteiden nojalla kyetä yksilöimään ja toimittamaan riittävät tiedot rekisteröidylle. Haasteita aiheuttaa se, että tietosuoja-asetuksessa ei avata tarkemmin termiä ”käsittelyyn liittyvä logiikka”, jolloin tietosuoja-asetuksen kannalta tarkasteltuna vaikuttaisi tulkinnanvaraiselta, kattaako termi esimerkiksi järjestelmän opetusdatan, tekniset tiedot hyödynnetyistä mallista tai päätöksentekosäännöt.

WP29 tietosuojatyöryhmä on tarkentanut rekisterinpitäjälle asetettuja tiedon toimittamista koskevia velvoitteita ja suosittaa rekisterinpitäjää toimittamaan tietoa siitä, miten profiili on luotu.²⁸⁰ Toisaalta tämäkin jättää avoimeksi sen, voiko rekisterinpitäjä toimia suosituksen mukaisesti toimittamalla kuvauksen profiiliin luomiseen käytettävästä datasta kuitenkin toimittamatta varsinaista dataa. Syrjivyyden tunnistamisen kannalta nimittäin saattaa olla merkitystä sillä, toimitetaanko rekisteröidylle varsinainen data vai rekisterinpitäjän sanoittama kuvaus datasta, joka saattaa joissakin tilanteissa olla liian epätarkka tai muutoin puutteellinen.²⁸¹

Hallinnon automaattista päätöksentekoa koskevaa yleislainsäädäntöä valmisteleavan työryhmän hallinnon automaattista päätöksentekoa, käyttöalaa ja läpinäkyvyyttä koskevissa säännösluonnoksissa todetaan, että toistaiseksi ei ole selvää, miten ”*merkitykselliset tiedot käsittelyyn liittyvästä logiikasta*” tulisi ymmärtää automaattisten hallintopäätösten kontekstissa. Säännösluonnoksissa viitataan oikeuskanslerin ratkaisukäytäntöön, jossa

²⁷⁹ Ks. läpinäkyvyyden osa-alueista 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 4.

²⁸⁰ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 33.

²⁸¹ Ks. tietojen ymmärrettävyydestä sekä yksinkertaisesta kielestä 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 7–11.

päätöksentekosääntöjä ja algoritmeja on pidetty merkityksellisinä tietoina automaattisessa päätöksenteossa.²⁸²

Toisaalta algoritmien osalta on huomattava myös WP29 tietosuojatyöryhmän suuntaviivat, joiden mukaan informointivelvoitteiden toteuttaminen ei välttämättä edellytä monimutkaisen selityksen toimittamista päätöksenteossa käytettävistä algoritmeista tai koko algoritmin paljastamista.²⁸³ Algoritmisten päätöksentekojärjestelmien osalta tulee huomata, että ainakin tiettyjen niihin liittyvien tietojen toimittamisen rekisteröidylle saattavat estää muiden osapuolten oikeudet esimerkiksi niiden sisältämään lähdekoodiin.²⁸⁴

Tietosuoja-asetuksen mukaisen etupainotteisen informoinnin fokus vaikuttaisikin olevan järjestelmää koskevan teknisen datan toimittamisen sijaan siinä, että rekisteröity ymmärtää päätöksenteon merkittävyyden ja sen seuraukset itselleen. Edellä kuvattujen oikeuskanslerin ja WP29 tietosuojatyöryhmän kannanottojen perusteella konkreettiseksi toimitettavaksi tiedoksi voidaan tunnistaa päätöksentekosäännöt, algoritmit ja tieto siitä, miten profiili on luotu.

Rekisterinpitäjän on mahdollista toimittaa edellä kuvattujen tietojen lisäksi teknisempiä tietoja kuten opetusdataa rekisteröidylle, mutta tällöin rekisterinpitäjän tulee kiinnittää erityistä huomiota siihen, ettei rekisteröityä kuormiteta liialla tiedolla ja että tiedot ovat helposti ymmärrettävässä muodossa, jolloin kohdeyleisön keskimääräisen edustajan tulee kyetä ymmärtämään tiedot.²⁸⁵

On kuitenkin olennaista havaita, että koneoppivissa järjestelmissä algoritmit oppivat aiempien korrelaatioiden ja päätelmien perusteella. Algoritmit voivat myös oppia aiemman perusteella siten, että ne voivat määrittää, mitkä aiempien ennustavien algoritmien mittaamiseen käytetyt mittaristot olivat tehokkaita ja suositella uusia iteraatioita testaukseen.²⁸⁶ Tästä taas seuraa se, että ennen tietyn päätöksen tekemistä koskevaan

²⁸² Hallinnon automaattista päätöksentekoa koskevaa yleislainsäädäntöä valmistelevan työryhmän muistio 2021, s. 7. Säännösluonnoksissa kuitenkin huomautetaan, ettei EU-lainsäädäntöä ole mahdollista selittää tai tulkita kansallisessa lainsäädännössä, mistä seuraa, että tietosuoja-asetuksen informointivelvoitteen tulkinta ja soveltaminen jää päätöksentekojärjestelmää hyödyntävän viranomaisen tehtäväksi. Ks. Hallinnon automaattista päätöksentekoa koskevaa yleislainsäädäntöä valmistelevan työryhmän muistio 2021, s. 7.

²⁸³ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 27.

²⁸⁴ Hakkarainen ym. 2020, s. 28–30.

²⁸⁵ 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 7. On huomattava, että jokainen algoritmi on erilainen ja jokaista algoritmia tulee arvioida sen toimintaympäristön kannalta, jotta on mahdollista määrittää sen ymmärtämistä edellyttävät tiedot. Diakopoulos on sitä mieltä, että tämä prosessi on sekä tekninen, että ihmiskeskeinen ja edellyttää tiedot vastaanottavan yleisön tietoa koskevien tarpeiden huomioimista. Ks. Diakopoulos 2016, s. 61.

²⁸⁶ Citron – Pasquale 2014, s. 5.

ajanjaksoon ulottuva läpinäkyvyys ei välttämättä kuvasta päätöksenteon hetkellä järjestelmässä olevaa syrjivyyttä.²⁸⁷

Toisaalta on kysyttävä, missä määrin rekisterinpitäjänkään on mahdollista toimittaa tietoa tekemättömän päätöksen seurauksista, vaikkakin tietosuoja-asetuksessa viitataan tässä yhteydessä mahdollisiin seurauksiin. Kyseeseen voisi tulla tilanne, jossa rekisterinpitäjä kertoisi algoritmisesta päätöksenteosta mahdollisesti seuraavan sen, että rekisteröidylle ei tulla myöntämään haettua etuutta, mikäli hakija ei täytä tiettyjä kriteereitä. Kyseeseen ei tällöin kuitenkaan tule se, että rekisterinpitäjä kertoisi (tai pystyisi välttämättä) kertomaan, että etuutta ei myönnetä siitä syystä, että algoritmi tulee todennäköisesti luokittelemaan hakijan tiettyyn luokkaan tietyn vinoutuneen datan tai muun epäasiallisen prosessin tuloksena.

4.5.2. Päätöksenteon jälkeen annettavat tiedot

Algoritmisen päätöksenteon syrjivyyden havaitsemisen kannalta olennainen vaihe kohdistuu aikaan, jolloin yksittäinen päätös on tehty. Syy tälle on se, että algoritmisen järjestelmä on tässä kohtaa päätöksenteon elinkaarta soveltanut relevantteja rekisteröityä ja tarvittavia seikkoja koskevia korrelaatioita ja päätenyt näiden pohjalta lopputulokseen, eli päätökseen. Yksittäistä henkilöä koskevaa syrjintää on itseasiassa tapahtunut vasta tässä kohtaa päätöksentekoprosessia.²⁸⁸

Edellä kuvattu informaation ajallinen ulottuvuus ilmenee myös tietosuoja-asetuksessa. Tietosuoja-asetuksen 13–14 artiklojen ja 15 artiklan eroavaisuudet nousevat esille suhteessa rekisteröidyn oikeuteen tietää päätöksenteon logiikka. Nimittäin tietosuoja-asetuksen artiklat 13(2)(f) ja 14(2)(g) mahdollistavat ainoastaan yleisen tiedon antamisen järjestelmien toiminnasta, koska ne velvoittavat rekisterinpitäjää antamaan tiedot käsittelyn alkaessa, jolloin mitään päätöstä ei ole vielä tehty. Toisaalta tietosuoja-asetuksen 15(1)(h) artikla saattaa mahdollistaa tiettyä päätöstä koskevan tiedon toimittamisen rekisteröidylle, koska sen mukaan rekisteröidylle tulee antaa pääsy häntä koskeviin henkilötietoihin (henkilötietoja on tällöin siis jo käsitelty ja päätös mahdollisesti tehty).²⁸⁹

²⁸⁷ Malgieri – Comandé 2017, s. 256.

²⁸⁸ Oma kysymyksensä on se, onko mahdollista pitävällä tavalla argumentoida, että syrjintää on tapahtunut jo laajemmin ymmärrettyä siinä kohtaa, kun järjestelmään on luotu mahdollisuus soveltaa syrjintään johtavia parametreja kyseiseen henkilöön. Käytännössä tosin henkilölle ei vielä edellä mainitussa vaiheessa ole aiheutunut konkreettisia kielteisiä seurauksia.

²⁸⁹ Malgieri – Comandé 2017, s. 255.

Malgierin ja Comandén mukaan rekisterinpitäjien tulisi tietosuoja-asetuksen 15(1)(h) artiklan mukaisesti kertoa automaattisen päätöksenteon varsinaisesta logiikasta, eikä pelkkä yleinen automaattisen päätöksentekojärjestelmän toimintaa koskeva tieto ole riittävää.²⁹⁰

Vaikka tietosuoja-asetuksen 15(1)(h) artiklan muotoilu on identtinen tietosuoja-asetuksen 13(2)(f) artiklan ja 14(2)(g) artiklan kanssa, on WP29 tietosuojatyöryhmä tulkinnut sen sisällön laajemmaksi verrattuna kyseisiin artikloihin. Tämä ilmenee siitä, että WP29 tietosuojatyöryhmä nostaa esille seuraavat tietotyypit, jotka rekisteröidylle tulisi toimittaa 15 artiklan perusteella:

- Päätöksentekoprosessissa huomioidut tekijät ja tieto siitä, miten niitä on painotettu kokonaistasolla.²⁹¹
- Profiilin luomiseen syöttötietoina käytetyt tiedot sekä tieto profiilista ja niistä segmenteistä, joihin rekisteröity on sijoitettu.²⁹²

WP29 tietosuojatyöryhmä edellyttää 15 artiklan yhteydessä merkittävän määrän tietoa toimittamista rekisteröidylle hieman yllättäenkin huomioiden WP29 tietosuojatyöryhmän kannan siitä, ettei rekisteröityä tulisi kuormittaa liiallisella tiedolla.²⁹³ Rekisteröidylle toimitettu suuri määrä tietoa saattaa nimittäin johtaa siihen, ettei läpinäkyvyys toteudu ideaalilla tavalla, vaan tärkeät tiedot jäävät rekisteröidyltä havaitsematta tietomassan seasta.²⁹⁴

On kuitenkin huomattava, että tietosuoja-asetuksen 15 artikla edellyttää sitä, että rekisteröity pyytää rekisterinpitäjää toimittamaan nämä tiedot. Tällöin rekisteröity ei saa kyseisiä tietoja, mikäli tämä ei osaa pyytää niitä rekisterinpitäjältä. Tähän taas vaikuttaa osaltaan se, miten rekisterinpitäjä on informoinut rekisteröityä päätöksenteosta ennen henkilötietojen käsittelyn aloittamista tietosuoja-asetuksen 13 ja 14 artiklojen perusteella. Lisäksi rekisterinpitäjän tulee kertoa rekisteröidylle tämän oikeudesta saada pääsy tietoihinsa.²⁹⁵ Toisaalta on huomattava, että kielteisen päätöksen saaneella rekisteröidyllä on todennäköisesti intressi pyytää jonkinlaisia lisätietoja päätöstään koskien.

²⁹⁰ Malgieri – Comandé 2017, s. 256.

²⁹¹ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 29.

²⁹² 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 18.

²⁹³ 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 7.

²⁹⁴ Stohl ym. 2016, s. 133.

²⁹⁵ Tietosuoja-asetuksen 13(2)(b) ja 14(2)(c) artiklojen mukaan rekisterinpitäjän on informoitava rekisteröityä tämän oikeudesta pyytää rekisterinpitäjältä pääsy häntä itseään koskeviin henkilötietoihin sekä oikeudesta pyytää kyseisten tietojen oikaisemista tai poistamista taikka käsittelyn rajoittamista tai vastustaa käsittelyä sekä oikeudesta siirtää tiedot järjestelmästä toiseen.

Keskusteltaessa konkreettisen päätöksen ymmärrettävyydestä algoritmisen päätöksenteon kontekstissa on arvioitava myös sitä, miten hallintolain mukainen päätösten perusteluvollisuus suhteutuu kokonaisuuteen. Tätä asiaa käsiteltiin ylempänä luvussa 4.4.

4.5.3. Johtopäätökset algoritmista päätöksentekoa koskevien tietojen toimittamisesta suhteessa algoritmisen päätöksenteon elinkaareen

Kuten edellä kuvattiin, on algoritmisten päätöksentekojärjestelmien läpinäkyvyys hahmotettavissa monitasoiseksi ja moniulotteiseksi kokonaisuudeksi kattaen päätöksentekoprosessin eri vaiheet ja järjestelmien eri osat.²⁹⁶ Keskusteltaessa algoritmisten päätöksentekojärjestelmien ymmärrettävyydestä onkin tärkeää tunnistaa tapauskohtainen rekisteröidyn intressi saada tietoa päätöksestä.²⁹⁷

Olenainen säädös algoritmisen päätöksenteon läpinäkyvyyteen liittyen on tietosuoja-asetus, sillä siinä asetetaan rekisterinpitäjälle kattavat velvoitteet informoida rekisteröityä automaattisesta päätöksenteosta sekä etupainotteisesti ennen yksittäisen päätöksen tekemistä, että rekisteröidyn käyttäessään oikeuksiaan myös käsittelyn myöhemmässä vaiheessa.²⁹⁸ Lisäksi hallintolain 45 §:n mukainen päätösten perusteluvollisuus kattaa konkreettisen päätöksen tekemisen jälkeisen tietojen toimittamisen omalta osaltaan.

WP29 tietosuojatyöryhmän mukaan etenkin tietojen käsittelyn ollessa monimutkaista, teknistä tai odottamatonta, tulee rekisterinpitäjän tietosuoja-asetuksen 13 ja 14 artiklan mukaisten tietojen lisäksi selittää rekisteröidylle erikseen ja selkeää kieltä käyttäen henkilötietojen käsittelyn tärkeimmät seuraukset (miten henkilötietojen käsittely vaikuttaa rekisteröityyn).²⁹⁹

Nähdäkseni algoritmisen päätöksenteon syrjivyyden tunnistettavuutta koskevassa analyysissä paljastuu kuitenkin olennainen haaste ajatukselle siitä, että päätöksenteon syrjivyyden voidaan kommunikoida rekisteröidylle tavalla, joka antaa hänelle mahdollisuuden

²⁹⁶ Ks. Hakkarainen ym. 2020, s. 4. Algoritmisen päätöksenteon avoimuus ja läpinäkyvyys on lopulta melko monisyinen asia hallinnon kontekstissa kuten ilmenee osuvasti oikeuskanslerin Kelaa koskevassa ratkaisussa (Valtioneuvoston oikeuskansleri, oikeuskanslerin ratkaisu 2021, OKV/131/70/2020 20.4.2021), jossa oikeuskansleri totesi, että hallintolaki ja tietosuoja-asetus asettavat samat lähtökohdat läpinäkyvyydelle ja avoimuudelle, jotka pitävät sisällään rekisteröidyn tarpeista lähtevän avoimen ja ymmärrettävästi esitettävän tiedon rekisteröidyn oikeuksiin ja velvollisuuksiin vaikuttavasta henkilötietojen käsittelystä eli päätöksenteosta. Asiakkaan oikeuksien turvaaminen edellyttää tietosuoja-asetuksen edellyttämällä tavalla tietoa automaattisesta päätöksenteosta. Myös hallintolain mukaisesti perusteltu päätös on edellytys oikeuksien turvaamiselle. Jotta päätös täyttäisi tietosuoja-asetuksen läpinäkyvyyden vaatimukset muun muassa suojaomien osalta ja olisi hallintolain mukaisesti kielelliseltä asultaan ja rakenteellisesti selkeä ja ymmärrettävä sekä perusteltu hallintopäätös, tulisi automaattista päätöksentekoa koskevan tiedon ja automaattisessa päätöksenteossa hyödynnettyjen päätöksentekosääntöjen muodostaa osan päätöskokonaisuutta.

²⁹⁷ Ks. Hakkarainen ym. 2020, s. 19.

²⁹⁸ Ks. Tietosuoja-asetuksen 12–15 artiklat.

²⁹⁹ 29 Artiklan mukainen tietosuojatyöryhmä, WP260, 2017, s. 7.

ehkäistä syrjinnän toteutuminen. Nimittäin ennen kuin järjestelmä on soveltanut painotuksia tiettyihin syötteisiin ja luonut tarvittavat vasteet sisältävän päätöksen, ei ole vielä olemassa mitään konkreettista päätöstä, joka johtaisi syrjintään. Tällöin rekisteröidylle ei tosiasiassa voida kommunikoida ennaltaehkäisevästi juuri häntä koskevan päätöksen mahdollista syrjivyyttä. Päätöksen tekemisen jälkeen taas ollaan tilanteessa, jossa rekisteröidylle on jo mahdollisesti syntynyt haitallisia seurauksia päätöksestä.³⁰⁰

Tällöin ei voida puhua tosiasiallisesta vaikutusmahdollisuudesta, vaan näennäisestä oikeudesta valvoa päätöksentekoa, johon liittyvästä syrjivyydestä rekisteröidyn ei välttämättä ole mahdollista muodostaa riittävää ja oikea-aikaista ymmärrystä.³⁰¹

Toisaalta rekisteröidyn kannalta oikeudenkäynnistä hallintoasioissa annetun lain 3 luvun mukainen päätöksen valitusoikeus nousee relevantiksi oikeussuojakeinoksi hallintopäätösten kontekstissa, jolloin rekisteröidyn ei voida katsoa jäävän kuitenkaan vaille oikeussuojakeinoja.

Rekisterinpitäjien tulee edellä kuvatun perusteella antaa riittävät tiedot algoritmisen päätöksenteon yksityiskohdista rekisteröidylle. Toimitettavien tietojen tarkkaa sisältöä ei ole kuitenkaan yksilöity tietosuojasetuksessa, jolloin rekisterinpitäjien tulee tapauskohtaisesti arvioida tarvittavien tietojen sisältöä huomioiden algoritmisen päätöksenteon elinkaari.

Tässä luvussa kuvatut toimitettavat tiedot eivät kuitenkaan sellaisenaan takaa sitä, että rekisteröity kykenisi tulemaan tietoiseksi mahdollisesta algoritmisen päätöksenteon syrjivyydestä. Olennaista syrjivyyden tunnistamiseksi on toimitettujen tietojen sisällön ymmärrettävyys rekisteröidyn kannalta. Seuraavassa luvussa tarkastellaan haasteita algoritmisen päätöksenteon tunnistettavuuden osalta nimenomaan rekisteröityjen ymmärryksen asettamien rajoitteiden kautta.

³⁰⁰ Ks. tietosuojasetuksen informaation ajallisesta ulottuvuudesta Malgieri – Comandé 2017, s. 255–256.

³⁰¹ Edwards ja Veale esittävät, että huolimatta koneoppimista koskevien selitysten potentiaalisen ilmenemisestä erinäisissä tutkimuksissa, on olemassa riski siitä, että selitysoikeuden edellyttäminen olisi harhaanjohtavaa tai voisi johtaa jopa niin sanottuun läpinäkyvyyden harhaan. He vertaavat läpinäkyvyyden harhaa tietosuojasetuksen suostumuksen yhteydessä esitettyyn niin sanottuun merkityksettömään suostumukseen. He nostavat selitysoikeuden sijasta potentiaalisiksi algoritmisten järjestelmien hallinnan välineiksi muita tietosuojasetuksen mukaisia rekisteröidyn oikeuksia kuten oikeuden poistaa tiedot ja oikeuden siirtää tiedot järjestelmästä toiseen, mutta korostavat, että nämäkin oikeudet ovat liiaksi riippuvaisia yksilön oikeuksista, kun kyseeseen tulee ryhmiä koskeva haitallinen vaikutus. Ks. Edwards-Veale 2017, s. 81.

5. HAASTEET ALGORITMISEN PÄÄTÖKSENTEON SYRJIVYYDEN TUNNISTAMISESSA

5.1. Rekisteröidyn kyvyttömyys ymmärtää algoritmista päätöksentekoa koskevaa tietoa

Kuten edellä kuvattiin, ei algoritmisen päätöksentekojärjestelmän ymmärrettävyyttä sen mahdollisesti sisältämän syrjivyyden tunnistamiseksi tulisi käsittää staattisena asiana, vaan ymmärrettävyyden yhteydessä on sen sijaan tarkasteltava laajemminkin sitä, mitä rekisteröity haluaa ymmärtää päätöksenteosta, eli mitkä hänen tiedolliset tarpeensa kussakin tilanteessa ovat.³⁰²

Vaikka organisaatio saattaisi relevantin tiedon sitä pyytävien saataville, ei tämä sellaisenaan takaa sitä, että rekisteröidyt kykenevät ymmärtämään kyseisen tiedon merkityksen päätöksentekoprosessin syrjivyyttä koskien. Rekisteröityjen tulisikin ymmärtää vastaanottamansa päätöksentekoa koskeva tieto, jotta tiedon voidaan katsoa olevan saatavilla heille.³⁰³

Edellä tarkasteltiin sitä, millaisia mekanismeja lainsäädännön kautta on pyritty luomaan rekisteröidyn ymmärryksen lisäämiseksi algoritmisesta päätöksenteosta. Erityisesti edellä kuvattu tietosuoja-asetuksen läpinäkyvyyden periaatteesta johdettu selitysoikeus ja tähän yhdistetty päätösten perusteluvollisuus pyrkivät nähdäkseni nostamaan rekisteröidyn ymmärryksen tasolle, jossa hän on tietoinen algoritmisen päätöksenteon olemassaolosta sekä yksittäisen päätöksen perusteluista ja siihen vaikuttaneista syistä.

On kuitenkin kysyttävä, onko tietosuoja-asetuksen läpinäkyvyyden ja hallintolain päätösten perusteluvollisuuden mukaisilla tiedoilla tosiasiasa mahdollista selittää algoritmisen päätöksenteon sisältö rekisteröidylle tavalla, jolla hänen on mahdollista tunnistaa järjestelmän sisältämä piilevä syrjivyy?

Tässä kohtaa on tarpeen tarkastella ylätasolla algoritmisen päätöksenteon syrjivyyden tunnistamisen haasteiden juurisyyt. Kuten aiemmin on kuvattu, algoritmien toiminta perustuu siihen, että ne käyttävät syötteenä dataa, jonka perusteella ne tuottavat vasteen. Algoritmien läpinäkymättömyys ilmenee siten, että luokittelujen kohteina olevilla

³⁰² Hakkarainen ym. 2020, s. 19.

³⁰³ Stohl ym. 2016, s. 129. On myös huomattava, että edes järjestelmien kehittäjät eivät aina välttämättä kykene selittämään monimutkaisten järjestelmien toimintaa. Ks. Ananny – Crawford 2018, s. 981.

henkilöillä on harvoin konkreettista ymmärrystä siitä, miten tai miksi tiettyyn luokitteluun on päädytty tietyn syötteen perusteella.³⁰⁴

Karkeasti yleistäen on mahdollista jaotella algoritmisten päätöksentekojärjestelmien syrjivyyden seuraaviin kategorioihin:

- 1) Syrjivyyden, joka muodostuu siitä, että päätöksentekojärjestelmä on ohjelmoitu ottamaan huomioon tiettyjä syrjiviä tekijöitä.
- 2) Syrjivyyden, joka muodostuu siitä, että järjestelmän opetusdatan mukana järjestelmään on päätyneet vinoumia, jotka johtavat syrjiviin lopputuloksiin.
- 3) Syrjivyyden, joka muodostuu siitä, että koneoppimiseen perustuva algoritmisen päätöksentekojärjestelmä opettaa itseään aiemmin tekemiensä vinoumia sisältävien päätösten pohjalta.³⁰⁵

Ensimmäisessä kohdassa mainittu syrjivyyden voisi ilmetä esimerkiksi siten, että järjestelmään on määritelty seuraava ”sääntö”: jos hakija on X sukupuolta = lopputulos Y, joka poikkeaa toisen sukupuolen osalta tehtävästä lopputuloksesta tavalla, joka on syrjivä.³⁰⁶

Toisen kohdan syrjivyyden aiheuttavat järjestelmän opetusdatassa olevat vinoumat. Erinäisessä datassa ilmenevät tyypillisesti yhteiskunnassa vallitsevat tai hiljattain vallinneet yhdenvertaisuutta koskevat ongelmat. Esimerkiksi tietyn sukupuolen pienempi osuus jonkin ammatin harjoittajista ilmenisi mahdollisesti siten, että järjestelmä oppisi tulkitsemaan, että tiettyä ammattia harjoittavat todennäköisemmin tietyn sukupuolen edustajat. Tästä taas voisi seurata syrjivää päätöksentekoa, mikäli järjestelmä soveltaa oppimaansa tavalla, joka ei huomioi oikeudenmukaisella tavalla käsillä olevaa tapausta. Toisaalta järjestelmää voidaan opettaa datalla, jonka edustavuus on puutteellista, jolloin opetusdata ei edes peilaa yhteiskunnassa vallitsevaa tai vallinnutta tilannetta.³⁰⁷

Kolmannen kohdan syrjivyyden saa alkunsa siitä, että koneoppivissa järjestelmissä algoritmit luovat uusia oppimismalleja aiempien korrelaatioiden ja päätelmien perusteella. Algoritmit voivat myös oppia aiemman perusteella siten, että ne voivat määrittää, mitkä aiempien

³⁰⁴ Edwards-Veale 2017, s. 25.

³⁰⁵ Barocas – Selbst 2016, s. 674–675 ja Citron – Pasquale 2014, s. 5–6 ja Kroll ym. 2017, s. 680–682. Kohdassa 1) tulee huomata, että ihmisen ohjelmoidessa päätöksentekosäännöt, tulee kyseeseen sääntöpohjainen automaatio, eikä koneoppiminen tältä osin. Ks. oikeusministeriön arviomuistio 2020, s. 11.

³⁰⁶ Oikeusministeriön arviomuistio 2020, s. 11–12.

³⁰⁷ Kroll ym. 2017, s. 680–682.

ennustavien algoritmien mittaamiseen käytetyt mittaristot olivat tehokkaita ja suositella uusia iteraatioita testaukseen.³⁰⁸

Seuraavaksi tarkastellaan tarkemmin järjestelmän osien, eli järjestelmien kehitysvaiheen datan, mallien, koodin sekä järjestelmien soveltamien korrelaatioiden ymmärrettävyyden haasteita suhteessa syrjivyyden tunnistamiseen.

5.1.1. Päätöksentekojärjestelmän kehitysvaiheen tiedot

Päätöksentekojärjestelmän algoritmien kehitysvaiheen data kuten opetusdata ja testausdata ovat yksi mahdollinen tapa pyrkiä toteuttamaan algoritmisen päätöksentekojärjestelmän läpinäkyvyys ja ymmärrettävyys, mutta tällöin on huomattava koneoppimisalgoritmien muuttumisen tuomat haasteet tiedon ajantasaisuudelle.³⁰⁹

Järjestelmän kehittämiseen hyödynnetyn datan osalta rekisteröidyn tulisi tietojen pohjalta kyetä arvioimaan useita eri tekijöitä kuten sitä, onko algoritmin datasta havaitsema kaava tarpeeksi täydellinen, sisältääkö se vinoumia, puuttuuko siitä tärkeitä kenttiä, onko data tarpeeksi korkealaatuista ja onko eri lähteistä peräisin oleva data integroitu tehokkaasti ja yhdenmukaisesti.³¹⁰ Käytännössä rekisteröidyn tulisi perehtyä mahdollisesti huomattavan suureenkin määrään dataa, josta hänen tulisi kyetä arvioimaan edellä kuvattuja tekijöitä.

Lisäksi valinnat päätöksenteossa huomioitavista datamalleista, eli niin sanottu ominaisuuksien valinta voi osoittautua ongelmalliseksi syrjinnän tunnistettavuuden kannalta. Kroll ym. ovat kuvanneet kolmenlaisia syrjivyyden kannalta ongelmallisia syötteitä koskevia valintoja:

- 1) Suojellun ryhmän jäsenyyden käyttäminen suoraan syötteenä (esim. sukupuolen huomioon ottava päätösprosessi).
- 2) Tietyn suojellun ryhmän arviointi yhtä suurella tarkkuudella toisten ryhmien kanssa, mutta riittämättömän kattavalla määrällä tekijöitä verrattuna toisten ryhmien arviointiin käytettäviin tekijöihin (esimerkiksi jos tiettyä ryhmää koskien ei ole riittävästi aikaisempaa dataa historiallisista syistä, voi päätöksenteossa käytettävä data olla tämän ryhmän osalta vähemmän luotettavaa).

³⁰⁸ Citron – Pasquale 2014, s. 5.

³⁰⁹ Ananny – Crawford 2018, s. 982.

³¹⁰ O’Hara 2020, s. 5.

- 3) Sellaisten tekijöiden hyödyntäminen päätöksenteossa, jotka liittyvät tiettyyn ryhmään johtaen siihen, että tekijä kuvaa todellisuudessa päätöksenteossa kyseistä ryhmää.³¹¹

Tämän lisäksi haastetta ymmärrettävyydelle tuo se, että algoritmit hyödyntävät laajasti ja monimutkaisilla tavoilla eri tekijöitä päätöksenteossaan.³¹²

Rekisteröidyn voisi teoriassa olla mahdollista tunnistaa riittäväällä tavalla sellainen syrjivyyden, joka olisi havaittavissa suojellun ryhmän jäsenyyden käyttämisenä suoraan syötteenä (esimerkiksi sukupuolen syrjivällä tavalla huomioon ottava päätösprosessi). Nähdäkseni tällainen tilanne oli käsillä yhdenvertaisuus- ja tasa-arvolautakunnan päätöksessä koskien luottokelpoisuuden arviointimenettelyä.³¹³ Tällaisessa tilanteessa rekisteröidyn saatua tietoa päätöksen perustan muodostaneista seikoista (tieto siitä, että päätökseen vaikuttivat esimerkiksi hakijan sukupuoli, ikä tai kieli), saattaa hänelle muodostua epäily mahdollisesta syrjivyydestä näiden tekijöiden perusteella yhdistettynä kielteiseen päätökseen.

Toisaalta on verraten epätodennäköistä, että rekisteröity osaisi toimitetun datan perusteella tulkita päätöksen perustuneen riittämättömän kattavaan määrään tekijöitä (esimerkiksi jos rekisteröityä koskevaa ryhmää koskien ei ole riittävästi aikaisempaa dataa, voi päätöksenteossa käytettävä data olla tämän ryhmän osalta vähemmän luotettavaa) tai sellaisten tekijöiden hyödyntämiseen päätöksenteossa, jotka käytännössä johtavat siihen, että tekijä kuvaa todellisuudessa ryhmää, johon rekisteröity kuuluu. Tämä edellyttäisi pääsyä vertailudataan ja kykyä hahmottaa saatavilla olevan datan määrä ja laatu.³¹⁴

On myös huomattava, että vaikka tietylle syrjintäperusteelle ei annettaisi tekoälyjärjestelmässä merkitystä ja tämä tieto poistettaisiin datasta, voi tekoäly tästä huolimatta saada esille datasta sellaista henkilöä koskevaa tietoa, jolla on vahva liityntä syrjintäperusteeseen.³¹⁵

³¹¹ Kroll ym. 2017, s. 681.

³¹² Edwards - Veale 2017, s. 50.

³¹³ Kyseisessä tapauksessa luottolaitosyhtiö oli evännyt hakijalta laskuluoton myöntämisen luoton hakijan asuinpaikan ja tämän sukupuolen, iän ja kielen perusteella. Yhdenvertaisuus- ja tasa-arvolautakunnan päätös, 21.3.2018, 216/2017.

³¹⁴ Stohl ym. 2016, s. 129–131.

³¹⁵ Yhdenvertaisuusvaltuutettu, tekoäly ja yhdenvertaisuus. <<https://syrjinta.fi/tekoaly#syrjintakiellon-valvonta>> Etusivu>Syrjintä ja yhdenvertaisuus>Aiheet>Tekoäly (haettu 2.4.2022). Myös luotonannon yhteydessä on esitetty, ettei sukupuolen poistaminen päätöksenteon muuttujista vaikuttaisi merkittävästi mallien ennustusten tehokkuuteen. Ks. Andreeva – Matuszyk 2019, s. 22.

On siis mahdollista, että yhdistelemällä henkilöön liittyvää tietoa, tekoäly muodostaa itsenäisesti välillisesti syrjiviä ennusteita ja johtopäätöksiä ilman sen laatijan tai sitä käyttävän tahon tahtoa.³¹⁶

Koneoppimisen osalta lähdekoodin informatiivisuutta vähentää myös se, että päätöksentekosäännöt syntyvät automaattisesti järjestelmän suorittaman data-analyysin tuloksena tavalla, jota ihmismielen ei ole mahdollista kaikissa tilanteissa selittää. Tällaisessa tapauksessa lähdekoodi itsessään paljastaa vain vähän päätöksenteon arviointia suorittavalle taholle, sillä koodi näyttää vain sovelletun koneoppimismenetelmän, mutta ei datapainotteista päätöksentekosääntöä.³¹⁷ Hakkaraisen ym. mukaan järjestelmien toimintaa havainnollistavien keinojen ja toisaalta lähdekoodin tason toteutuksen välille muodostuukin eräänlainen kuilu, jonka ylittämiseksi tulee toteuttaa tulkintaa.³¹⁸

5.1.2. Järjestelmien soveltamat korrelaatiot ja piilevä syrjivyyys

Tässä alaluvussa kuvataan sitä, miten haastavaa on havaita niin sanottua piilevää syrjivyyttä, joka aiheutuu siitä, että koneoppimisjärjestelmien algoritmien päätökset perustuvat syy-seuraussuhteen sijaan korrelaatiosuhteiden havaitsemiseen.³¹⁹

Käytännössä järjestelmien painotukset ja vasteet eivät usein ole merkityksellisiä ja ymmärrettäviä yksittäiselle henkilölle, koska järjestelmä pystyy havaitsemaan kuvioita ja säännönmukaisuuksia tietoaaineistoista tavalla, johon ihmismieli ei käytännössä kykene.³²⁰

Tämän havainnollistamiseksi voidaan kuvitella tilanne, jossa rekisteröidylle toimitettaisiin tiedot häntä koskevassa päätöksessä käytetyistä tekijöistä (syötteet). Tietojen pohjalta rekisteröidylle muodostuu käsitys, ettei päätöksenteossa ole huomioitu syrjivyyteen viittaavia seikkoja (esimerkiksi hakijan sukupuolta, ikää tai kieltä), vaan päätökseen ovat vaikuttaneet esimerkiksi rekisteröidyn siviilisääty, asuinalue ja harrastukset. Nämä tiedot eivät ainakaan sellaisenaan vaikuttaisi viittaavaan ainakaan suoraan syrjivyyteen (toki hakija voisi kyseenalaistaa näiden relevanssin), mutta koneoppimisjärjestelmä on saattanut korrelaatiosuhteiden kautta tulkita tekijöitä tavalla, joka johtaa käytännössä hakijan syrjintään. Arvioidakseen esimerkiksi sitä, onko postinumeroiden ja henkilön alkuperän välillä korrelaatiota tietoaaineistossa, rekisteröidyn tulisi saada tieto joko siitä, että

³¹⁶ Yhdenvertaisuusvaltuutettu, tekoäly ja yhdenvertaisuus. <<https://syrjinta.fi/tekoaly#syrjintakiellon-valvonta>> Etusivu>Syrjintä ja yhdenvertaisuus>Aiheet>Tekoäly (haettu 3.4.2022).

³¹⁷ Kroll ym. 2017, s. 638.

³¹⁸ Ks. Hakkarainen ym. 2020, s. 24.

³¹⁹ Selbst – Barocas 2018, s. 1089-1090.

³²⁰ O'Hara 2020, s. 1.

päätöksenteon tietoa sisältää alkuperää koskevan attribuutin tai rekisteröidyn tulee saada tietoa kyseisen alueen väestörakenteesta.³²¹ Käytännössä korrelaatio-suhteisiin perustuvasta päätöksenteosta seuraa se, että rekisteröidyn ei ole hänelle mahdollisesti toimitettavan tiedonkaan perusteella mahdollista merkityksellisellä tavalla arvioida sitä, kohdistuuko häneen piilevää syrjintää.

Esimerkkinä algoritmisen päätöksenteon ymmärrettävyyden haasteista syrjivyyden kontekstissa voidaan nostaa esille ns. COMPAS-tapaus. Hakkaraisen ym. mukaan tapauksessa abstraktin tason esitys ja ymmärtäminen ei ollut riittävää päätöksentekojärjestelmän kokonaisvaltaiseen tarkasteluun, sillä piilotettujen vinoutumien paikantaminen vaatisi järjestelmän konepellin alle kurkistamista, eli perehtymistä järjestelmän opetusmateriaalina käyttämään dataan ja järjestelmän tekemisiin päätöksiin.³²²

Kuitenkaan COMPAS-tapauksessa kuvattu konepellin alle kurkistaminen ei tuo rekisteröidylle riittävää ymmärrystä piilevän syrjivyyden tunnistamiseksi, koska kahden asian välisen korrelaation päättelyprosessin kuvailu ei vielä itsessään vastaa kysymykseen siitä, miksi kyseiseen korrelaatioon perustuva päätös oli oikeutettu, eli toisin sanoen se ei vastaa kysymykseen korrelaation aiheuttajasta. Kuvitellaan esimerkiksi tilanne, jossa punaisiin housuihin pukeutuvalta henkilöltä evätään autovakuutus. Päätöksentekoa kuvaava selitys voi sisältää kuvauksen siitä, miten vinoumista vapaa tietoa sisältävä paljasti huomattavan korrelaation auto-onnettomuuksien ja punaisten housujen käyttämisen välillä. Toisaalta yksittäisen rekisteröidyn osalta selitys voi aiheuttaa hämmennystä siitä, miksi punaisten housujen käyttäminen korreloi liikenneonnettomuuksien kanssa ja voisiko hän vaihtamalla housujensa väriä saada vakuutuksen vai tulisiko hänen tehdä jotakin muuta vaikuttaakseen itselleen edullisen päätöksen syntyyn.³²³

5.1.3. Johtopäätökset rekisteröidyn kyvyttömyydestä ymmärtää algoritmista päätöksentekoa koskevaa tietoa

Tässä luvussa kuvatun perusteella on mahdollista todeta, että algoritmisen päätöksenteon syrjivyyden voi saada alkunsa useasta eri tekijästä. Näitä tekijöitä ovat esimerkiksi järjestelmän kehitysvaiheen data ja syötteitä koskevat valinnat.³²⁴

³²¹ Veale –Binns 2017, s. 4.

³²² Hakkarainen ym. 2020, s. 31.

³²³ O'Hara 2020, s. 4.

³²⁴ O'Hara 2020, s. 5 ja Kroll ym. 2017, s. 681.

Burrellin mukaan algoritmisia päätöksiä onkin mahdollista selittää teoriassa esimerkiksi kuvailemalla syitä, miksi tietty luokittelupäätös tehtiin. Tällöin kyseeseen tulee kuitenkin ihmismielen suorittaman tulkitsevan päättelyn soveltaminen tilastollisen optimoinnin matemaattiseen prosessiin, jolloin selitykseen jää epävarmuuksia kuten tiettyjen sanojen painotusten merkitys luokitteluun. Tästä taas seuraa se, ettei tällainen ”miksi” kysymykseen vastaava selitys välttämättä ole oikeanlainen.³²⁵

Erityisesti yllä kuvattu korrelaatio-suhteiden hyödyntämiseen perustuva niin sanottu piilevä syrjintä vaikuttaisi sellaiselta, jota rekisteröidyn ei käytännössä ole mahdollista tunnistaa tälle toimitetusta järjestelmää koskevasta datasta tai selityksestä.³²⁶ Tästä johtuen rekisterinpitäjien tulisi tunnistaa tällaiset järjestelmät ja käyttötapaukset erityisen riskialttiina ja etupainotteisesti varmistua niiden syrjimättömyydestä.

Barretin mukaan läpinäkyvyys ei olekaan ihmeratkaisu, mutta siitä voi olla hyötyä kaikista räikeimpien järjestelmätason ongelmien aisoihin laittamiseen. Etenkin läpinäkyvyydestä voi olla hyötyä tarkoituksellisen tai osittain tarkoituksellisen syrjinnän ehkäisemisessä kuten ohjelmoijan suorittamien muuttujia koskevien valintojen tekemisessä, datan keräämisen metodeissa ja automaatiovinoumia sisältävän tai edistävän päätöksentekojärjestelmien käytön suhteen. Läpinäkyvyys ei kuitenkaan poista Barretin mukaan dataan sisältyvän tahattoman järjestelmätason syrjinnän vaikutuksia.³²⁷

Vaikka rekisteröidyn ei olisi mahdollista tunnistaa toimitetuista tiedoista etenkin korrelaatio-suhteiden luomaa piilevää syrjivyyttä, tulee rekisterinpitäjän silti varmistaa päätöksenteon ja henkilötietojen käsittelyn läpinäkyvyyden ja ymmärrettävyyden toteutuminen rekisteröidyn suuntaan. Tämä kattaa erityisesti päätöksenteon merkittävyyden ja potentiaaliset seuraukset sekä henkilötietojen käsittelyä koskevat tiedot.³²⁸ Tällä tavoin rekisteröity voi valvoa henkilötietojensa käsittelyä (esimerkiksi jos päätös perustuu virheelliseen tai puutteelliseen rekisteröityä koskevaan dataan tai rekisteröity kokee tulleensa muutoin kohdelluksi väärin).³²⁹

Wachter ym. ovat esittäneet, että algoritmisten päätösten selitysten ei tarvitse olla riippuvaisia siitä, ymmärtävätkö ihmiset algoritmisten päätöksentekojärjestelmien

³²⁵ Burrell 2016, s. 9.

³²⁶ O’Hara 2020, s. 1 ja s. 4.

³²⁷ Barrett 2017, s. 344.

³²⁸ Ks. Tietosuoja-asetuksen 13(2)(f), 14(2)(g) ja 15(1)(h) artikla.

³²⁹ Rekisteröidyn henkilötietojen käsittelyyn liittyvistä oikeuksista säädetään tietosuoja-asetuksessa. Tietosuoja-asetuksen 77 artiklan mukaan rekisteröity voi tehdä valituksen valvontaviranomaiselle, mikäli rekisteröity katsoo, että häntä koskevien henkilötietojen käsittelyssä rikotaan tietosuoja-asetusta.

toimintaa. Heidän mukaansa järjestelmien tulkittavuus on sinänsä tavoittelemisen arvoinen asia, mutta selitykset on mahdollista toimittaa avaamatta 'mustaa laatikkoa'. Selitysten tarkoituksena olisi tällöin edesauttaa rekisteröityjä toimimaan tietyllä tavalla pelkän ymmärryksen edistämisen sijaan.³³⁰ Nähdäkseni selitysten tarkoitusta rekisteröidyn toimintaa ohjaavana työkaluna on kuitenkin vaikeaa yhteensovittaa syrjivään päätöksentekoon, sillä rekisteröity ei voi estää ominaisuuksiinsa perustuvaa syrjintää toimimalla toisin.³³¹

Viime vuosina on myös noussut esille enenevässä määrin ajatuksia sellaisten tekoälyohjelmien kehittämiseen, joiden toimintaperiaatteet on mahdollista selittää ja joiden päättelyketjut ovat ymmärrettävissä (niin sanottu explainable AI, XAI). Hakkaraisen ym. mukaan XAI on ennen kaikkea tarkoitettu asiantuntija-avoimuuden lisäämiseen, sillä vain asiantuntijoilla on käytännössä mahdollisuus ymmärtää järjestelmiin liittyvää koodia. Asiantuntijoiden rooliksi voisikin muodostua tulkin rooli tekoälyjärjestelmien ja niiden käyttäjille välille.³³²

On myös ehdotettu, että algoritmisten päätöksentekojärjestelmien toimintaa voitaisiin valvoa siten, että rekisteröidyille tarjottaisiin mahdollisuus saada varmistus siitä, että heihin kohdistui samanlainen prosessi kuin muihin, samoja periaatteita noudatetaan jokaisen päätöksen osalta, päätös voidaan toisintaa ja että päätöksentekoprosessin periaatteet oli määritelty ennen kuin kyseinen rekisteröity oli järjestelmän tiedossa.³³³ Tässäkin toteutustavassa haasteita aiheuttanee järjestelmässä mahdollisesti vallitseva piilevä syrjivyyys.

³³⁰ Wachter ym. 2018, s. 843.

³³¹ Tällainen rekisteröidyn toimintaa ohjaileva selitys voisi olla hyödyllinen tilanteissa, joissa rekisteröity voi todella toimimalla toisin vaikuttaa päätöksen sisältöön (esimerkiksi muuttamalla ajotyyliään). Ks. 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 28.

³³² Ks. Hakkarainen ym. 2020, s. 18. Tällaiseen rooliin voitaisiin nähdäkseni pohtia esimerkiksi niin sanottuja organisaatioiden sisällä toimivia tekoälyvastaavia tietosuojavastaavien tapaan.

³³³ Goldenfein 2019, s. 51.

6. JOHTOPÄÄTÖKSET

Tässä tutkielmassa on tunnistettu eri syrjivyyden ilmenemismuotoja koneoppimiseen perustuvissa algoritmisissa päätöksentekojärjestelmissä. Algoritmisen päätöksenteon syrjivyyks voi ilmetä järjestelmän elinkaaren eri vaiheissa. Se voi saada alkunsa jo järjestelmän kehitysvaiheessa järjestelmään päätyneiden vinoumien tai puutteiden takia.³³⁴ Myös järjestelmän syötteitä koskevat valinnat voivat luoda syrjivyyttä.³³⁵ Lisäksi järjestelmien mahdollisesti toteuttama korrelaatio-suhteisiin perustuva luokittelu voi johtaa piilevään syrjintään.³³⁶

Algoritmisten järjestelmien syrjivyydestä saattaa aiheutua kielteisiä seurauksia niiden kohteena oleville henkilöille.³³⁷ Hallinnon kontekstissa kyseeseen tulevat syrjinnän perusteella syntyneet kielteiset päätökset kuten päätökset, joissa hakijalle myönnettäisiin jokin etuus huonommilla ehdoilla tai se evättäisiin kokonaan syrjivällä tavalla.

Edellä kuvatut seuraukset saattavat vaikuttaa epätodennäköiseltä, mutta maailmalta on saatavilla esimerkkejä, jossa tekoälyn hyödyntäminen esimerkiksi rikoksen uusimisriskin arvioinnissa johti syrjintään. Lisäksi myös Suomessa tekoälyn hyödyntäminen hallinnossa on aiheuttanut valvontaviranomaisten toimenpiteitä, joskaan nimenomaan syrjivyyden osalta tässä tutkielmassa ei havaittu viranomaiskäytäntöä.³³⁸

Jotta rekisteröity voisi valvoa oikeuksiaan ja riitauttaa häntä koskevan päätöksen merkityksellisellä tavalla, tulisi rekisteröidyn saada riittävät tiedot päätöksestä ja siihen vaikuttaneista tekijöistä. Läpinäkyvyyden käsite nousee tässä kontekstissa olennaiseksi, sillä sen kautta algoritmisia järjestelmiä hyödyntävien organisaatioiden tulee kertoa päätöksenteosta rekisteröidyille. Hallinnon kontekstissa päätösten perusteluvelvollisuus pyrkii varmistamaan asianosaisen mahdollisuuden valvoa päätöksen oikeellisuutta ja turvautua oikeussuojakeinoihin.³³⁹

Analyysin tuloksena havaittiin, että tietosuoja-asetuksen läpinäkyvyyden ja hallintolain mukaisen päätösten perusteluvelvollisuuden mukaiset päätöksenteon ymmärrettävyyttä

³³⁴ Euroopan parlamentin tutkimus 2019, s. 7.

³³⁵ Kröll ym. 2017, s. 681.

³³⁶ Citron – Pasquale 2014, s. 5.

³³⁷ 29 Artiklan mukainen tietosuojatyöryhmä, WP251, 2017, s. 5–6.

³³⁸ Ks. COMPAS-tapauksesta Hakkarainen ym. 2020, s. 16 ja hallinnon kontekstissa eduskunnan oikeusasiamies, Verohallinnon automatisoitu päätöksentekomenettely ei täytä perustuslain vaatimuksia - Tiedotteet, 2019 <<https://www.oikeusasiamies.fi/fi/-/verohallinnon-automatisoitu-paatoksentekomenettely-ei-tayta-perustuslain-vaatimuksia>> (haettu 11.3.2022).

³³⁹ Hakkarainen ym. 2020, s. 18.

edistävät rekisterinpitäjän velvoitteet ovat osittain tulkinnanvaraisia siltä osin, miten tarkasti rekisterinpitäjän tulee kertoa algoritmisen järjestelmän toiminnasta.³⁴⁰

Tutkielmassa kuitenkin tunnistettiin eräitä tietotyyppisiä, joita rekisterinpitäjien tulisi toimittaa rekisteröidylle. Tietotyypit vaihtelevat sen osalta, tuleeko kyseeseen ennen päätöstä vai päätöksen jälkeen toimitettavat tiedot.³⁴¹

Tutkielmassa pohdittiin, onko pelkkä läpinäkyvyys riittävää, jotta rekisteröity voi tunnistaa syrjivän päätöksenteon. Vaikuttaisi siltä, että erityisesti monimutkaisten piilevää syrjivyyttä sisältävien järjestelmien osalta läpinäkyvyys sellaisenaan ei riitä varmistamaan riittävää ymmärrystä, jotta rekisteröity voi tunnistaa syrjivyyden.³⁴²

Sen sijaan rekisteröidyn olisi teoriassa mahdollista havaita syrjivyyteen viittaavien ominaisuuksien hyödyntäminen päätöksenteon kriteereinä. Myös puutteellisen opetusdatan käyttäminen voisi teoriassa olla havaittavissa rekisteröidyn toimesta, joskin tämä edellyttäisi todennäköisesti työlästä prosessia.³⁴³

Rekisteröidyn ymmärrystä edesauttavaksi on akateemisessa keskustelussa nostettu tietosuoja-asetuksesta mahdollisesti johdettavissa oleva selitysoikeus. Selitysoikeuden sisältöä ei ole mahdollista määrittää yksiselitteisesti, mutta sen on esitetty kattavan tiedot siitä, miten ja miksi jokin päätös tehtiin.³⁴⁴

Tämän tutkielman olennainen havainto on se, että vaikka rekisteröidylle toimitettaisiin tietoa ja selityksiä algoritmista päätöksenteosta, ei tämä itsessään takaa sitä, että rekisteröity kykenisi ymmärtämään kyseistä tietoa tavalla, jonka kautta hän voi tunnistaa piilevän syrjivyyden. Tämä johtuu koneoppimisjärjestelmien perimmäisestä toimintalogiikasta: järjestelmät pystyvät havaitsemaan kuvioita ja säännönmukaisuuksia tietoineistoista tavalla, johon ihmismieli ei käytännössä kykene.³⁴⁵

Tämä ei kuitenkaan merkitse sitä, etteikö läpinäkyvyydellä ja selityksillä olisi tärkeää roolia rekisteröidyn henkilötietojen käsittelyn ja myös yleisemmällä tasolla algoritmisen päätöksenteon ymmärrettävyyden kuten sen merkittävyyden ja mahdollisten seurausten saattamisella rekisteröidyn tietoon. Läpinäkyvyydellä ja selityksillä varmistetaan osaltaan

³⁴⁰ Ks. tietosuoja-asetuksen läpinäkyvyyttä koskevien artiklojen tulkinnanvaraisuudesta tässä kontekstissa Brkan 2018, s. 1 ja perusteluvelvollisuuden sisällöstä algoritmisen päätöksenteon yhteydessä Hakkarainen ym. 2020.

³⁴¹ Malgieri – Comandé 2017, s. 255–256.

³⁴² Ks. läpinäkyvyyden illuusiosta Hakkarainen ym. 2020, s. 8.

³⁴³ Kroll ym. 2017, s. 681 ja O'Hara 2020, s. 5.

³⁴⁴ Roscher ym. 2020, s. 42203.

³⁴⁵ O'Hara 2020, s.1.

se, että rekisteröity voi valvoa henkilötietojensa käsittelyä ja riitauttaa päätöksen siltä osin, kuin kyseeseen tulee muu kuin piilevä syrjintä.³⁴⁶

Tässä tutkielmassa tarkasteltiin lisäksi sitä, miten hyvin läpinäkyvyys laajempina periaatteena sisältäen hallinnon avoimuuden, onnistuu vastaamaan algoritmisen päätöksenteon mukanaan tuomiin haasteisiin. Johtopäätöksenä tähän voidaan todeta, että tietosuoja-asetuksen ja päätösten perusteluvollisuuden kautta toteutettava läpinäkyvyys ja avoimuus eivät kaikissa tilanteissa ole riittäviä keinoja, jotta rekisteröity voi tulla tietoiseksi häntä koskevasta syrjinnästä päätöksenteossa. Tämä koskee erityisesti tilanteita, joissa syrjintä aiheutuu koneoppimisjärjestelmässä vallitsevasta piilevästä syrjivyydestä. Tällöin voidaan havaita, että teknologinen kehitys on johtanut tilanteeseen, jossa aiemmat normit eivät välttämättä pysty vastaamaan kehityksen mukanaan tuomiin haasteisiin sellaisenaan.

Hallinnon algoritmisen päätöksenteon kehittämisen osalta erityishuomion tulisikin keskittyä piilevän syrjivyyden tapausten etupainotteiseen tunnistamiseen, riskien minimointiin ja tarvittaessa järjestelmän käyttöönotosta luopumiseen, ellei riittäviä keinoja minimoida syrjivyyttä ole mahdollista ottaa käyttöön. Päätöksentekojärjestelmien syrjivyyttä tulisikin nähdä ensisijaisesti ehkäistä muilla keinoin kuin rekisteröidyn suorittaman valvonnan toimesta.³⁴⁷

Eryteisesti hallinnon kontekstissa on tunnistettava algoritmisen päätöksenteon yhteydessä hyödynnettäviin eräisiin tekoälyjärjestelmiin liittyvä korostunut riski siitä, että ne sisältävät piilevää syrjintää, jota rekisteröityjen ei ole mahdollista tunnistaa riippumatta siitä, miten paljon ja minkä sisältöistä tietoa heille toimitetaan päätöksentekoon liittyen. Mikäli hallinnon kontekstissa halutaan hyödyntää esimerkiksi koneoppimista päätöksenteossa, on tarpeen käydä periaatteellisen tason keskustelu siitä, miten pitkälle rekisteröidyn ymmärryksen toteutumisen ideaalista ollaan valmiita joustamaan. Lopulta kyse on luottamuksesta: rekisteröityjen tulee voida luottaa viranomaisiin ja viranomaisten tulee varmistua, että ne voivat luottaa hyödyntämiensä tekoälyjärjestelmien asianmukaiseen toimintaan.³⁴⁸

³⁴⁶ Ks. läpinäkyvyydestä automaattisessa päätöksenteossa tietosuoja-asetuksen 13(2)(f), 14(2)(g) ja 15(1)(h) artikla ja läpinäkyvyyden tavoitteista Hakkarainen ym. 2020, s. 18–19.

³⁴⁷ Goldenfein 2019, s. 41. Esimerkiksi havaitut vinoumat tulee poistaa ja heikkolaatuisen datan käyttö lopettaa. Ks. Barrett 2017, s. 344.

³⁴⁸ Hakkarainen ym. 2020, s. 22.