



HELSINGIN YLIOPISTO

# Käsitys ihmisen ja järjestelmän vuorovaikutuksesta tekoälyssäädöksessä

Haasteet oikeudellistuvassa vuorovaikutuskäsityksessä ja teknologiasäätelyssä

Oikeustieteen maisterin koulutusohjelma  
Tekoälyoikeus Maisterintutkielma

Laatija:  
Hilma Laamanen

Ohjaajat:  
EU-oikeuden dosentti Suvi Sankari  
Apulaisprofessori (tekoälyn yhteiskunnallisoikeudelliset vaikutukset) Riikka Koulu

23.4.2024  
Helsinki

**Tiedekunta:** Oikeustieteellinen tiedekunta

**Koulutusohjelma:** Oikeustieteen maisterin tutkinto

**Opintosuunta:** Tekoälyoikeus: Tekoälyn yhteiskunnallisoikeudelliset vaikutukset

**Tekijä:** Hilma Laamanen

**Työn nimi:** Käsitys ihmisen ja järjestelmän vuorovaikutuksesta tekoälysäädöksessä – Haasteet oikeudellistuvassa vuorovaikutuskäsityksessä ja teknologiasääntelyssä

**Työn laji:** Maisterintutkielma

**Kuukausi ja vuosi:** Huhtikuu, 2024

**Sivumäärä:** 78 + XIV

**Avainsanat:** Ihminen-kone vuorovaikutus, ihminen-järjestelmä vuorovaikutus, yhteiskunnallinen oikeustutkimus, empiirinen oikeustutkimus, HCI-tutkimus, HABA-MABA, Ihminen-kone symbioosi, EU-oikeus, teknologiasääntely, oikeuden digitalisaatio, tekoälyasetus, tekoälysäädös, tekoälyjärjestelmä, ihmisvalvonta

**Ohjaaja tai ohjaajat:** Suvi Sankari ja Riikka Koulu

**Säilytyspaikka:** Helsingin yliopiston kirjasto

**Tiivistelmä:** Tekoälyteknologioiden käyttöönotto yhteiskunnan eri osa-alueilla luo mahdollisuuksia, mutta aiheuttaa myös riskejä. Tämän vuoksi EU:n tekoälysäädöksellä luodaan horisontaaliset ja yhdenmukaistetut säännöt tekoälyjärjestelmien kehittämiseksi ja käytölle. EU:n pyrkimyksenä on sääntelyn avulla edistää tekoälyteknologian käyttöönottoa ja varmistaa samalla unionin oikeushyvien toteutuminen. Tekoälyteknologian käyttöönoton hidasteena on teknologiaa kohtaan koetun luottamuksen puute. Luottamuksen varmistamiseksi sääntelykehys perustuu seitsemälle keskeiselle vaatimukselle, joista yksi on ihmisen toimijuus ja ihmisen suorittama valvonta. Vaatimuksen myötä ihmisen ja tekoälyjärjestelmän vuorovaikutus oikeudellistuu tekoälysäädöksessä.

Tutkielman tavoitteena on selvittää oikeudellistuvan vuorovaikutuskäsityksen sisältö. Ihmisen ja järjestelmän vuorovaikutusta on tutkittu HCI-tutkimukseksi nimetyllä (eng. "Human-Computer Interaction Research") tieteenalalla 1960-luvulta alkaen. Tekoälysäädös analysoidaan teorialähtöisesti käyttäen analyysirunkoa, joka perustuu HCI-tutkimuksessa käytettyihin käsitteisiin ja mallinnuksiin ihminen-kone vuorovaikutuksesta. Tutkielman tiedonintressi on lainopillinen, sillä tavoitteena on empiiristä menetelmää hyödyntämällä selvittää lainsäädäntöön omaksutun vuorovaikutuksen tarkka sisältö. Tutkielmassa on myös sääntelyteoreettinen näkökulma, sillä HCI-tutkimuksen vuorovaikutuskäsitys mahdollistaa lainsäädännön laadun ja sääntelykeinojen kriittisen arvioinnin.

Oikeudellinen vuorovaikutuskäsitys muodostuu erityisesti ihmisvalvonnasta, ihmiseen palautettavasta toimijuudesta ja ihmisen suorittamasta päätöksenteosta. Oikeudellinen vuorovaikutuskäsitys on erittäin dikotominen ja vastaa HCI-tutkimuksessa automaation aikakaudella kehitettyjä vuorovaikutuksen käsitteellistämisen tapoja. Oikeudellinen vuorovaikutuskäsitys ei ole ongelmaton, sillä HCI-tutkimuksen näkökulmasta se ei vastaa teknologian kehityksen astetta. Oikeudellinen lähestymistapa luottamuksen rakentamiseen perustuu HCI-tutkimuksessa tehottomiksi tai epäsuotuisiksi todetuille keinoille.

Teknologia ja sen sääntely aiheuttaa yhteiskunnalle ja oikeudelle tiettyjä haasteita. Oikeudellista vuorovaikutuskäsitystä selittää oikeuden antroposentrismi. Digitalisaation aiheuttama haaste oikeuden antroposentrismille näkyy tekoälysäädöksessä erityisesti ristiriitana ihmisen vuorovaikutuksessa saaman roolin ja järjestelmien ympäristössään aiheuttamien tosiasiallisten vaikutusten välillä. Ristiriitaisuus on uhka sääntelyn tarkoituksenmukaisuudelle ja siten sääntelyn legitimiydelle.

# Sisällysluettelo

<b>Lähteet</b>	<b>I</b>
<b>Virallislähteet</b>	<b>I</b>
Suomi	I
EU	I
EU:n sekundäärioikeus	I
Komission ehdotukset	II
Muut EU:n virallislähteet	II
Kansainväliset sopimukset	III
<b>Kirjallisuus</b>	<b>III</b>
<b>Verkkolähteet</b>	<b>XII</b>
<b>Suulliset tiedonannot</b>	<b>XIV</b>
<b>1 Johdanto</b>	<b>1</b>
<b>1.1 Tausta ja merkitys</b>	<b>1</b>
<b>1.2 Teoreettinen viitekehys, tutkimuskysymykset &amp; metodit</b>	<b>2</b>
1.2.1 Teoreettinen viitekehys	2
1.2.2 Tutkimuskysymykset ja tutkielman rajaukset	4
1.2.3 Metodit	6
<b>1.3 Aineiston kuvailu &amp; tutkielman rakenne</b>	<b>8</b>
1.3.1 Aineisto	8
1.3.2 Tekoälysäädöksen tausta – sääntelyn tarve	9
1.3.3 Tekoälysäädös EU-lainsäädännön systematiikassa	11
1.3.4 Tekoälysäädöksen sisäinen systematiikka	12
1.3.5 Rakenne	13
<b>2 HCI-tutkimus</b>	<b>14</b>
<b>2.1 Tutkimusperinteestä ja historiasta</b>	<b>14</b>
2.1.1 Tutkimusperinteen synty	14
2.1.2 HCI-tutkimus vuorovaikutuksen tutkimuksen historiana	15
<b>2.2 Vuorovaikutuksen käsitteellistetyt mallit</b>	<b>16</b>
2.2.1 HABA-MABA (“Humans are better at & Machines are better at”)	16
2.2.2 Man-Machine Symbiosis	18
<b>2.3 Malleihin ja vuorovaikutukseen olennaisesti liittyvät käsitteet</b>	<b>20</b>
2.3.1 Automaatio	20
2.3.2 Autonomia	22
2.3.3 Toimijuus HCI-tutkimuksessa	24
<b>2.4 Jakson lopuksi</b>	<b>26</b>
<b>3 Aineiston analyysi</b>	<b>28</b>
<b>3.1 Kuvaus analyysirungosta- ja menetelmistä</b>	<b>28</b>

<b>3.2</b>	<b>Dikotomisen vuorovaikutuskäsityksen ilmentyminen</b>	<b>29</b>
3.2.1	Tarkkarajainen ja staattinen roolitus	29
3.2.2	Ihmisvalvonnan vaatimus luo jyrkän hierarkkian	31
3.2.3	Järjestelmien toimijuus uhkana ihmisen toimijuudelle	33
3.2.4	Ihmisen asettamat erinäiset tavoitteet	35
<b>3.3</b>	<b>Symbioottisen vuorovaikutuskäsityksen häivähdykset</b>	<b>37</b>
<b>3.4</b>	<b>Lopuksi</b>	<b>39</b>
<b>4</b>	<b>Oikeuden digitalisaatio ja teknologiasääntelyn haasteet</b>	<b>41</b>
<b>4.1</b>	<b>Oikeuden antroposentrismi ja oikeuden digitalisaatio</b>	<b>42</b>
4.1.1	Ihmiskeskeinen oikeus	42
4.1.2	Haasteiden käsitteellistäminen: Oikeuden digitalisaatio	43
<b>4.2</b>	<b>Teknologiasääntelyn haasteet</b>	<b>45</b>
4.2.1	Teknologiasääntelyn tutkimuksen suhde sääntelyyn	45
4.2.2	Teknologian käsitteestä	46
4.2.3	Sääntelyn määritelmä	48
4.2.4	Yleisen edun perustelut teknologian käyttöönotossa	50
4.2.5	Teknologian kaksoisrooli sääntelyssä	52
4.2.6	Teknologiasääntelyn tehokkuus ja kohdentuminen	53
4.2.7	Sääntelyvalintojen vaikutuksesta	56
<b>5</b>	<b>Oikeudellistuvan vuorovaikutuskäsityksen ydin ja ongelmat</b>	<b>57</b>
5.1.1	Ihmisvalvonta luotettavan tekoälyn edellytyksenä	57
5.1.2	Ihmisen toimijuus luotettavan tekoälyn edellytyksenä – Järjestelmä välineenä	62
5.1.3	Ihmisen päätöksenteko luotettavan tekoälyn edellytyksenä	70
5.1.4	Lopuksi	75
<b>6</b>	<b>Loppupäätelmät</b>	<b>77</b>

## Lähteet

### Virallislähteet

#### Suomi

HE 145/2022 vp Hallituksen esitys eduskunnalle julkisen hallinnon automaattista päätöksentekoa koskevaksi lainsäädännöksi

Oikeusministeriö. (2015). Syyttäjälaitoksen ja yleisten tuomioistuinten asian- ja dokumentinhallinnan kehittämishankkeen (AIPA) jatkotoimikausi 1.12.2021–31.12.2022 Saatavilla:

<https://oikeusministerio.fi/hanke?tunnus=OM007:00/2015>

Valtiovarainministeriö. (2023). Koneoppiminen digitaalisen turvallisuuden teknisessä valvonnassa. Valtiovarainministeriön julkaisuja, 2023:56.

Valtiovarainministeriö. Saatavilla: <http://urn.fi/URN:ISBN:978-952-367-641-1>

#### EU

##### EU:n sekundäärioikeus

Euroopan komissio. (2000). Komission asetus (EY) N:o 2082/2000, annettu 6 päivänä syyskuuta 2000, Eurocontrol-standardien hyväksymisestä sekä Eurocontrol-standardien hyväksymisestä annetun neuvoston direktiivin 93/65/ETY muuttamisesta annetun direktiivin 97/15/EY muuttamisesta. Official Journal of the European Union.

Euroopan komissio. (2014). Komission asetus (EU) N:o 1302/2014, annettu 18 päivänä marraskuuta 2014, Euroopan unionin rautatiejärjestelmän liikkuvan kaluston osajärjestelmää ”veturit ja henkilöliikenteen liikkuva kalusto” koskevasta yhteentoimivuuden teknisestä eritelmästä (ETA:n kannalta merkityksellinen teksti). Official Journal of the European Union.

Euroopan parlamentti & Euroopan unionin neuvosto. (2016). Euroopan parlamentin ja neuvoston direktiivi (EU) 2016/2102, annettu 26 päivänä lokakuuta 2016, julkisen sektorin elinten verkkosivustojen ja mobiilisovellusten saavutettavuudesta (ETA:n kannalta merkityksellinen teksti).

Euroopan parlamentti & neuvosto. (2016). 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 (yleinen tietosuojasäädös) (GDPR) (ETA:n kannalta merkityksellinen teksti).

Euroopan parlamentti & Euroopan unionin neuvosto. (2019). Euroopan parlamentin ja neuvoston direktiivi (EU) 2019/882, annettu 17 päivänä huhtikuuta 2019, tuotteiden ja palvelujen esteettömyysvaatimuksista (ETA:n kannalta merkityksellinen teksti).

Euroopan Parlamentti. (2024). Regulation on artificial intelligence (TA-9-2024-0138\_EN). Saatavilla [https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf)

### **Komission ehdotukset**

Euroopan komissio. (19. helmikuuta 2020). COM(2020) 66 final: Komission tiedonanto Euroopan parlamentille, neuvostolle, Euroopan talous- ja sosiaalikomitealle ja alueiden komitealle - Euroopan datastrategia.

Euroopan komissio. (21. huhtikuuta 2021). Ehdotus Euroopan parlamentin ja neuvoston asetukseksi tekoälyä koskevista yhdenmukaisista säännöistä (Tekoälysäädös) ja tiettyjen unionin säädösten muuttamisesta (COM(2021) 206 final, 2021/0106(COD)).

### **Muut EU:n virallislähteet**

Euroopan parlamentti. (16. helmikuuta 2017). Päätöslauselma suosituksista komissiolle robotiikkaa koskevista yksityisoikeudellisista säännöistä (2015/2103(INL)). Official Journal of the European Union.

Euroopan komissio. (6. tammikuuta 2024). Interinstitutional File: 2021/0106(COD): Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - Analysis of the final compromise text with a view to agreement.  
<https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf>

Euroopan komissio. (2020). Valkoinen kirja tekoälystä - Eurooppalainen lähestymistapa huipputaiteeseen ja luottamukseen COM(2020) 65 final.

Haettu osoitteesta [https://commission.europa.eu/system/files/2020-03/commission-white-paper-artificial-intelligence-feb2020\\_fi.pdf](https://commission.europa.eu/system/files/2020-03/commission-white-paper-artificial-intelligence-feb2020_fi.pdf)

Euroopan parlamentti. (2019). Luotettavaa tekoälyä koskevat eettiset ohjeet. Haettu osoitteesta [https://www.europarl.europa.eu/meetdocs/2014\\_2019/plmrep/COMMITTEE\\_ES/JURI/DV/2019/11-06/Ethics-guidelines-AI\\_FI.pdf](https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEE_ES/JURI/DV/2019/11-06/Ethics-guidelines-AI_FI.pdf)

### **Kansainväliset sopimukset**

Kansainvälinen työjärjestö (ILO). (1989). Yleissopimus alkuperäiskansojen ja heimokansojen oikeuksista riippumattomissa maissa (Nro 169)

Yhdistyneet kansakunnat. (2007). Yhdistyneiden kansakuntien julistus alkuperäiskansojen oikeuksista.

### **Kirjallisuus**

Aalto-Heinilä, M. (2009). Tahtoteoria ja intressiteoria oikeuksien funktion selittäjinä. *Oikeus*, 38(2), 138-158.

Aarnio, A. (2011). Luentoja lainopillisen tutkimuksen teoriasta. Helsingin yliopiston oikeustieteellinen tiedekunta.

Bainbridge, L. (1983). Ironies of Automation. *Automatica*, 19(6), 775-779.

Baldwin, R., Cave, M., & Lodge, M. (2011). *Understanding Regulation: Theory, Strategy, and Practice* (2nd ed.). Oxford University Press.  
<https://doi.org/10.1093/acprof:osobl/9780199576081.001.0001>

Bellantyne, N. (2019). Epistemic trespassing. *Mind*, 128(510), 367–395.  
<https://doi.org/10.1093/mind/fzx042>.

Bennett Moses, L. (2013). How to Think About Law, Regulation and Technology: Problems with 'Technology' as a Regulatory Target. *Law, Innovation and Technology*, 5(1), 1-20. UNSW Law Research Paper No. 2014-30.

- Black, J. (2002). Critical Reflections on Regulation. *Australian Journal of Legal Philosophy*, 27, 2-36.  
[https://www.researchgate.net/publication/265568938\\_Critical\\_Reflection\\_on\\_Regulation](https://www.researchgate.net/publication/265568938_Critical_Reflection_on_Regulation)
- Bødker, S. (2006). When second wave HCI meets third wave challenges. *NordiCHI '06: Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles*, 1–8. <https://doi.org/10.1145/1182475.1182476>.
- Boy, G.A. (2023). An Epistemological Approach to Human Systems Integration. *Technology in Society*, 74, 102298.
- Bradshaw, J. M., Feltovich, P. J., & Johnson, M. (2017). Human–Agent Interaction. In G. A. Boy (Ed.), *The Handbook of Human-Machine Interaction* (1st ed., pp. 283–300). CRC Press. <https://doi.org/10.1201/9781315557380-14>.
- Bradford, A. (2019). *The Brussels Effect: How the European Union Rules the World*. Oxford University Press.  
<https://doi.org/10.1093/oso/9780190088583.001.0001>
- Brownsword, R. (2008). *Rights, Regulation, and the Technological Revolution*. Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780199276806.001.0001>
- Brownsword, R., Scotford, E., & Yeung, K. (Eds.). (2017). *The Oxford Handbook of Law, Regulation, and Technology*. Oxford University Press. ISBN: 9780199680832.
- Cho, H., & Yoon, J. (2013). Toward a new design philosophy of HCI: Knowledge of collaborative action of “We” human-and-technology. In M. Kurosu (Ed.), *Human-Computer Interaction. Human-Centred Design Approaches, Methods, Tools, and Environments. HCI 2013 (Lecture Notes in Computer Science, vol. 8004, pp.32-40)*. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-39232-0\\_4](https://doi.org/10.1007/978-3-642-39232-0_4)
- Cohen, J. E. (2019). *Between Truth and Power: The Legal Constructions of Informational Capitalism*. Oxford University Press.  
<https://doi.org/10.1093/oso/9780190246693.001.0001>



- Cole, S. A., & Bertenthal, A. (2017). Science, Technology, Society, and Law. *Annual Review of Law and Social Science*, 13(1), 351–371.  
<https://doi.org/10.1146/annurev-lawsocsci-110316-113550>
- Constantino, J. (2022). Exploring Article 14 of the EU AI proposal: Accountability challenges of the human in the loop when supervising high-risk AI systems in public administration. Delft University of Technology.
- Corbridge, C., & Cook, C. A. (1997). The role of function allocation in the design of future naval systems. In E. Fallon, L. Bannon, & J. McCarthy (Eds.), *ALLFN'97 Revisiting the Allocation of Functions Issue: New Perspectives* (s. 73-88). IEA Press.
- Carroll, J. M. (2009). Human-Computer Interaction (HCI). Teoksessa M. Soegaard & R. F. Dam (Toimittajat), *Encyclopedia of Human-Computer Interaction. Interaktiivisen suunnittelun säätiö*, 21-62.  
[https://snoopedu.com/app/uploads/2022/03/Reading1\\_HCI.pdf](https://snoopedu.com/app/uploads/2022/03/Reading1_HCI.pdf)
- Crootof, R., Kaminski, M. E., & Price II, W. N. (2023). Humans in the loop. *Vanderbilt Law Review*, 76(2), 429.
- Curioni, A., Vesper, C., Knoblich, G., & Sebanz, N. (2019). Reciprocal information flow and role distribution support joint action coordination. *Cognition*, 187, 21-31. <https://doi.org/10.1016/j.cognition.2019.02.006>
- Dourish, P. (2004). *Where The Action Is: The Foundations of Embodied Interaction*. MIT Press.
- Enarsson, T., Enqvist, L., & Naarttijärvi, M. (2021). Approaching the human in the loop – legal perspectives on hybrid human/algorithmic decision-making in three contexts. *Information & Communications Technology Law*, 31(1), 123-153. <https://doi.org/10.1080/13600834.2021.1958860>
- Enqvist, L. (2023). ‘Human oversight’ in the EU artificial intelligence act: what, when and by whom? *Law, Innovation and Technology*, 15(2), 508-535.  
<https://doi.org/10.1080/17579961.2023.2245683>
- Ervasti, K. (1998). Eräitä näkökohtia empiirisen tiedon hyväksikäyttämisestä oikeustieteessä. *Lakimies*, 3, 364–388.

- Ervasti, K. T. (2017). *Lakimies, oikeus, yhteiskunta: oikeus yhteiskunnallisena käytäntönä*. Edita Publishing.
- Eteläpelto, A., Vähäsantanen, K., Hökkä, P. K., & Paloniemi, S. (2013). What is agency? Conceptualizing professional agency at work. *Educational Research Review*, 10, 45–65. <https://doi.org/10.1016/j.edurev.2013.05.001>
- Felt, U., Fouché, R., Miller, C. A., & Smith-Doerr, L. (Eds.). (2016). *The Handbook of Science and Technology Studies*. MIT Press.
- Fitts, P. M. (1951). Human Engineering for an Effective Air-Navigation and Traffic Control System. National Research Council, Division of Anthropology and Psychology, Committee on Aviation Psychology. <https://apps.dtic.mil/sti/citations/ADB815893>.
- Fuld, R. B. (2000). The fiction of function allocation, revisited. *International Journal of Human-Computer Studies*, 52(2), 217-233. <https://doi.org/10.1006/ijhc.1999.0286>
- Gondomar, R., & Mor, E. (2021). Understanding Agency in Human-Computer Interaction Design. In M. Kurosu (Ed.), *Human-Computer Interaction. Theory, Methods and Tools* (Vol. 12762, s. xxx-xxx). ISBN: 978-3-030-78461-4
- Grudin, J. (2012). A Moving Target—The Evolution of Human-Computer Interaction. Teoksessa J. Jacko (Toim.), *Human-Computer Interaction Handbook* (3. painos). Taylor & Francis.
- Hale, K. S., Stanney, K. M., & Schmorrow, D. D. (2007). Augmenting Cognition in HCI: Twenty-First Century Adaptive System Science and Technology. In Jacko, J. A. (Ed.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*.
- Hakkarainen, J. M., Koulu, R., & Markkanen, K. A. (2020). Läpinäkyvät algoritmit? Lähdekoodin julkisuus ja laillisuuskontrolli hallinnon digitalisaatiossa. *Edilex*, (2020/18). <https://www.edilex.fi/artikkelit/21042.pdf>
- Hancock, P. A. (2017). Imposing limits on autonomous systems. *Ergonomics*, 60\*(2), 284-291. <https://doi.org/10.1080/00140139.2016.1190035>

- Hildebrandt, M. (2019). Privacy As Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning. *Theoretical Inquiries in Law*, 19(1).
- Hildebrandt, M. (2020). Legal personhood for AI. Teoksessa M. Hildebrandt (Toim.), *Law for Computer Scientists and Other Folk* (s. 237–250). Oxford University Press. <https://doi.org/10.1093/oso/9780198860877.003.0009>
- Hirvonen, A. (2011). Mitkä metodit?: Opas oikeustieteen metodologiaan (Yleisen oikeustieteen julkaisuja; 17). Helsingin yliopisto, Oikeustieteellinen tiedekunta. <http://hdl.handle.net/10138/225264>
- Hoc, J.-M. (2000). From human–machine interaction to human–machine cooperation. *Ergonomics*, 43(7), 833-843. <https://doi.org/10.1080/001401300409044>
- Inga, J., Ruess, M., Robens, J. H., Nelius, T., Rothfuß, S., Kille, S., Dahlinger, P., Lindenmann, A., Thomaschke, R., Neumann, G., Matthiesen, S., Hohmann, S., & Kiesel, A. (2022). Human-machine symbiosis: A multivariate perspective for physically coupled human-machine systems. *International Journal of Human-Computer Studies*, 170, 102926. <https://doi.org/10.1016/j.ijhcs.2022.102926>
- Jenkins, M., Esemizie, A. O., Lee, V., & Mensingh, M. (2021). An investigation of “We” agency in co-operative joint actions. *Psychological Research*, 85(3), 1-15. <https://doi.org/10.1007/s00426-020-01462-6>
- Jones, M. L. (2017). Right to a human in the loop: Political constructions of computer automation & personhood from data banks to algorithms. *Social Studies of Science*, 47(2), 216-239. <https://doi.org/10.1177/0306312717692587>
- Koivisto, I. (2015). Oikeus on, miten se systematisoidaan? – Kysymys oikeudenalajaotuksesta ja hallinto-oikeudesta. *Lakimies*, 7–8, 954–972.
- Koops, B.-J. (2010). Ten Dimensions of Technology Regulation - Finding Your Bearings in the Research Space of an Emerging Discipline. In M.E.A. Goodwin et al. (Eds.), *Dimensions of Technology Regulation* (pp. 309-324). Nijmegen: WLP. Tilburg Law School Research Paper No. 015/2010. SSRN: <https://ssrn.com/abstract=1633985>
- Koulu, R. (2018). Digitalisaatio ja algoritmit – oikeustiede hukassa? *Lakimies*, 7–8, 840–867.

- Koulu, R. (2020). Human control over automation: EU policy and AI ethics. *European Journal of Legal Studies*, 12(1), 9-46.  
<https://doi.org/10.2924/EJLS.2019.019>
- Koulu, R. (2020). Proceduralizing control and discretion: Human oversight in artificial intelligence policy. *Maastricht Journal of European and Comparative Law*, 27(6), 720-735. <https://doi.org/10.1177/1023263X20978649>
- Koulu, R., & Alizadeh Westerling, F. (2023). Rethinking Access to Justice through digitalisation: User experiences of digital legal aid services. Paper presented at ILAG Conference 2023. <https://clp.law.harvard.edu/wp-content/uploads/2023/06/Session-9-Riikka-Koulo-and-Frida-Alizadeh-Westerling.pdf>
- Koulu, R. & Hakkarainen, J. M. (2023). Digitaalisten työvälineiden käyttö sovittelun laadun tukena. Oikeusministeriön julkaisuja, Selvityksiä ja ohjeita, Nro 2023:29. Oikeusministeriö, Helsinki.  
<https://julkaisut.valtioneuvosto.fi/handle/10024/165048>
- Koulu, R., Koivisto, I., & Larsson, S. (2024). User accounts: How technological concepts permeate public law through the AI Act. *Maastricht Journal of European and Comparative Law*.
- Koulu, R., Sankari, S., & Sormunen, S. K. (2022). Digitalisoituva julkishallinto: Käytettävyyttä kuuluu kaikille. *Edilex*, (2022/36).  
<https://www.edilex.fi/artikkelit/28209.pdf>
- Kuopus, J. (1988). Hallinnon lainalaisuus ja automatisoitu verohallinto: Oikeustieteellinen tutkimus kansalaisen oikeusturvasta teknistyvässä valtionhallinnossa. Lakimiesliiton Kustannus.
- Kurki, V. (2018). Ei vain oikeuskelpoisuutta – oikeussubjektikäsitteemme ongelmia ja uudelleenarviointia. *Lakimies*, 5, 469-492.
- Kurki, V. (2018). Voiko tekoäly olla oikeussubjekti? *Lakimies*, 7-8, 820-839.
- Kyriakou, K., & Otterbacher, J. (2023). In humans, we trust: Multidisciplinary perspectives on the requirements for human oversight in algorithmic

- processes. *Discover Artificial Intelligence*, 3, Article 44.  
<https://doi.org/10.1038/s44159-023-00123-0>
- Landry, S. J. (2013). Human computer interaction in aerospace. In J. A. Jacko (Ed.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (3rd ed., s. 778). New York: Taylor & Francis Group. <https://www-taylorfrancis-com.libproxy.helsinki.fi/pdfviewer/>
- Laux, J. (2023). Institutionalised distrust and human oversight of artificial intelligence: Towards a democratic design of AI governance under the European Union AI Act. *AI & Society*. <https://doi.org/10.1007/s00146-023-01675-4>
- Law, J. (2016). STS as Method. In U. Felt, R. Fouché, C. A. Miller, & L. Smith-Doerr (Eds.), *The Handbook of Science and Technology Studies*. MIT Press.
- Lehman-Wilzig, S. N. (1981). Frankenstein unbound: Towards a legal definition of artificial intelligence. *Futures*, 13(6), 442-457. [https://doi.org/10.1016/0016-3287\(81\)90100-2](https://doi.org/10.1016/0016-3287(81)90100-2)
- Lessig, L. (1999). *Code: And Other Laws of Cyberspace*. Basic Books.
- Licklider, J. C. R. (1960). Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics* HFE-1(1), 4–11.  
<https://doi.org/10.1109/THFE2.1960.4503259>.
- Loevinger, L. (1949). Jurimetrics: The Next Step Forward. *Minnesota Law Review*, 33(5), 455-493.
- Makkonen, K. (1981). Oikeudellisen ratkaisutoiminnan ongelmia: rakenneanalyttinen tutkimus. Suomalaisen Lakimiesyhdistyksen julkaisuja. B-sarja; n:o 191. Lainopillinen ylioppilastiedekunta. ISBN 951-9027-48-3.
- Malgieri, G. (2019). Automated decision-making in the EU Member States: The right to explanation and other “suitable safeguards” in the national legislations. *Computer Law & Security Review*, 35(5), Article 105327.  
<https://doi.org/10.1016/j.clsr.2019.05.002>
- Morse, S. J. (2017). Law, Responsibility, and the Sciences of the Brain/Mind. In R. Brownsword, E. Scotford, & K. Yeung (Eds.), *The Oxford Handbook of Law*,

Regulation and Technology. Oxford Handbooks. Oxford Academic.

<https://doi.org/10.1093/oxfordhb/9780199680832.013.7>

- Määttä, K. (2002). Oikeudellisen sääntelyn tutkimus – lastuja sääntelyteoriasta. *Oikeus*, 31(2), 132–142.
- Määttä, T. (2015). Metodinen pluralismi oikeustieteessä – ympäristöoikeudellisen tutkimuksen suuntauksat ja menetelmät. Teoksessa T. Miettinen (Toim.), *Oikeustieteellinen opinnäyte – Artikkeleita oikeustieteellisten opinnäytteiden vaatimuksista, metodista ja arvostelusta* (s. Edilex 2015/45). Edilex.
- Rogers, Y., Sharp, H., & Preece, J. (2011). *Interaction Design: Beyond Human-Computer Interaction* (3rd ed.). John Wiley & Sons.
- Roth, E. M., Sushereba, C., Militello, L. G., Diulio, J., & Ernst, K. (2019). Function Allocation Considerations in the Era of Human Autonomy Teaming. *Journal of Cognitive Engineering and Decision Making*, 13(4), 199–220.  
<https://doi.org/10.1177/1555343419878038>.
- Ruscheimer, H., & Hondrich, L. (2024). Automation bias in public administration - an interdisciplinary perspective from law and psychology. FernUniversität in Hagen.
- Sankari, S., Koulu, R., & Hakkarainen, J. M. (2022). Digitaalinen oikeus tutkimuksen kohteena. *Lakimies*, 120(1), 185-191.
- Schmidler, J., Knott, V., Hölzel, C., & Bengler, K. (2015). Human centered assistance applications for the working environment of the future. *Occupational Ergonomics*, 12(3), 83-95. <https://doi.org/10.3233/OER-150226>
- Shackelford, S. J., Asare, I. N., Dockery, R., Raymond, A., & Sergueeva, A. (2021). Should We Trust a Black Box to Safeguard Human Rights? A Comparative Analysis of AI Governance. *SSRN Electronic Journal*.  
<https://doi.org/10.2139/ssrn.3773198>
- Sheridan, T. B., & Ferrell, W. R. (1975). Man-Machine Systems: Information, Control and Decision Models of Human Performance. *Journal of Dynamic Systems Measurement and Control*, 27\*(1). DOI:10.2307/3009156

- Sheridan, T. B., & Parasuraman, R. (2005). Human-Automation Interaction. \*Reviews of Human Factors and Ergonomics, 1\*(1), 89-129. DOI: 10.1518/155723405783703082
- Simmler, M., & Frischknecht, R. (2021). A taxonomy of human-machine collaboration: capturing automation and technical autonomy. \*AI & Society, 36\*(3), 239-250. <https://doi.org/10.1007/s00146-020-01126-3>
- Smith, J. C. (1997). Machine intelligence and legal reasoning - The Charles Green Lecture in Law and Technology. *Chicago-Kent Law Review*, 73(1), Article 9.
- Solum, L. B. (1992). Legal personhood for artificial intelligences. *North Carolina Law Review*, 70(4), 1234-1240.
- Sterling, L. S., & Taveter, K. (2009). *The Art of Agent-Oriented Modeling*. Intelligent Robotics and Autonomous Agents series. The MIT Press. ISBN: 9780262013116
- Tarasti, L. (2002). Yhteiskunnan oikeudellistuminen. *Defensor Legis*, 4, 575-585.
- Thomas, B. (2011). A critical review of law's meaning of life: Philosophy, religion, Darwin and the legal person. *Journal of Law and Society*. <https://doi.org/10.1111/j.1467-6478.2011.00561.x>
- Tiensuu, P., Aalto-Heinilä, M., & Mäki-Petäjä-Leinonen, A. (toim.). (2023). *Itsemääräämisoikeus: Teoriasta käytäntöön*. Vastapaino. <https://doi.org/10.58181/VP9789523970069>
- Tuomi, J., & Sarajärvi, A. (2002). *Laadullinen tutkimus ja sisällönanalyysi*. Tammi.
- Tuori, K. (2003). *Oikeusjärjestys ja oikeudelliset käytännöt*. (Forum iuris). Helsingin yliopiston oikeustieteellinen tiedekunta. <https://doi.org/10.31885/9789515182630>
- Tuori, K. (2013). *Oikeusjärjestys ja oikeudelliset käytännöt*. Helsinki: Forum Iuris, Helsingin yliopiston oikeustieteellisen tiedekunnan julkaisut.
- Urquhart, L., & Rodden, T. (2016). *A Legal Turn in Human Computer Interaction: Towards Regulation by Design for the Internet of Things*. Social Science Research Network (SSRN).

- Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act: Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 4, 97-112.
- Viljanen, M. (2017). Algoritmien haaste – uuteen aineelliseen oikeuteen? *Lakimies*, 7–8, 1070–1087.
- de Visser, E. J., Pak, R., & Shaw, T. H. (2018). From ‘automation’ to ‘autonomy’: the importance of trust repair in human–machine interaction. *Ergonomics*, 61\*(11), 1409-1427. <https://doi.org/10.1080/00140139.2018.1457725>
- Wein, L. E. (1992). Responsibility of intelligent artifacts: toward an automation jurisprudence. *Harvard Journal of Law & Technology*, 6(1), 103-154. <https://jolt.law.harvard.edu/assets/articlePDFs/v06/o6HarvJLTech103.pdf>
- Wesche, J. S., & Sonderegger, A. (2019). When computers take the lead: The automation of leadership. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2019.07.027>
- Williams, R., & Edge, D. (1996). The social shaping of technology. *Research Policy*, 25(6), 865–899. [https://doi.org/10.1016/0048-7333\(96\)00885-2](https://doi.org/10.1016/0048-7333(96)00885-2)
- Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, 109(1), 121–136. <http://www.jstor.org/stable/20024652>
- Wright, P., Dearden, A., & Fields, B. (2000). Function Allocation: A Perspective from Studies of Work Practice. *International Journal of Human-Computer Studies*, 52(2), 335–355. <https://doi.org/10.1006/ijhc.1999.0292>.
- Žáčková, E. (2011). Man-computer symbiosis as a way of human cognitive enhancement. *Masaryk University Journal of Law and Technology*, 5(1), 133-144.

## Verkkolähteet

- Grady, P. (1. helmikuuta 2023). The AI Act Should Be Technology-Neutral. Center for Data Innovation. Saatavilla: <https://www2.datainnovation.org/2023-ai-act-technology-neutral.pdf> (viitattu 16.3.2024).



- Editors of Encyclopedia Britannica. (18. tammikuuta 2024). Personal computer. Encyclopedia Britannica. Saatavilla: <https://www.britannica.com/technology/personal-computer> (viitattu 27.1.2024.).
- Euroopan komissio. (23. syyskuuta 2021). Lausunto presidentti von der Leyenin puolesta yhteisessä lehdistötilaisuudessa presidentti Michelin kanssa G7-johdajien kokouksen jälkeen. Euroopan komissio - Lehdistötiedote. Saatavilla: [https://ec.europa.eu/commission/presscorner/detail/en/statement\\_23\\_6474](https://ec.europa.eu/commission/presscorner/detail/en/statement_23_6474) (viitattu 14.4.2024.).
- Euroopan komissio. (9. joulukuuta 2023). Statement by President von der Leyen on the political agreement on the EU AI Act. Euroopan komissio - Lehdistötiedote. Saatavilla: [https://ec.europa.eu/commission/presscorner/detail/en/statement\\_23\\_6474](https://ec.europa.eu/commission/presscorner/detail/en/statement_23_6474) (viitattu 17.2.2024.).
- Euroopan parlamentti. (6. joulukuuta 2023). Artificial Intelligence Act: Deal on comprehensive rules for trustworthy AI. Euroopan parlamentti - Lehdistötiedote. Saatavilla: <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai> (viitattu 14.4.2024.).
- European Parliament. (4. helmikuuta 2024). Regulation on artificial intelligence. Europe fit for the digital age - Legislative Train. Saatavilla: <https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence?sid=7801> (viitattu 27.3.2024.).
- Fehlinger, P. (12. lokakuuta 2023). Enabling the responsible use of technology at scale: Why Europe needs a regulatory technology innovation ecosystem. Sitra. Saatavilla: <https://www.sitra.fi/en/publications/enabling-the-responsible-use-of-technology-at-scale/#publication-content> (viitattu 16.3.2024.).
- IEEE. (2024). Autonomous Systems Industry Connections. Saatavilla: <https://standards.ieee.org/industry-connections/ec/autonomous-systems/> (viitattu 20.2.2024.).

IEEE. (13. joulukuuta 2016). Ethically Aligned Design Version 1 For Public Discussion. Saatavilla: [https://standards.ieee.org/wp-content/uploads/import/documents/other/ead\\_v2.pdf](https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf) (viitattu 20.2.2024).

IEEE. (14. tammikuu 2018). Ethically Aligned Design Version 2 For Public Discussion. Saatavilla: [https://standards.ieee.org/wp-content/uploads/import/documents/other/ead\\_v2.pdf](https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf) (viitattu 20.2.2024).

Open letter to the European Commission Artificial Intelligence and Robotics. Saatavilla: <https://robotics-openletter.eu/>. (viitattu 21.2.2024).

Stanford University. (2024). Artificial Intelligence Index Report 2024. Saatavilla: [https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI\\_AI-Index-Report-2024.pdf](https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_AI-Index-Report-2024.pdf) (viitattu 22.4.2024)

Von der Leyen, U. (18. heinäkuuta 2023). Speech at the International Conference on Industrial Informatics. European Commission. Saatavilla: [https://ec.europa.eu/commission/presscorner/detail/en/speech\\_23\\_3943](https://ec.europa.eu/commission/presscorner/detail/en/speech_23_3943) (viitattu 23.3.2024).

## **Suulliset tiedonannot**

Pyöreän pöydän keskustelu 20.9.2023: Keskustelijoina Maria Hauptmann (TEM), Tuomas Pöysti (Oikeuskansleri), Riikka Rosendahl (KKV)

# 1 Johdanto

## 1.1 Tausta ja merkitys

Vastauksena viimeaikaisimpaan datavetoisten ja vuorovaikutteisten teknologioiden esiinmarssiin Euroopan unioni tulee ensimmäisenä maailmassa kohdistamaan sääntelyä tekoälyjärjestelmiin.<sup>1</sup> Euroopan parlamentin ja neuvoston asetus tekoälyä koskevista yhdenmukaistetuista säännöistä (myöhemmin “tekoälysäädös”) on kunnianhimoinen yritys säännellä tekoälyjärjestelmiä teknologianeutraalisti ja horisontaalisesti.<sup>2</sup> Tekoälysäädös tulee määrittämään tekoälyjärjestelmien kehittämistä, markkinoille saattamista ja käyttöä unionissa. Säädöksen pyrkimyksenä on edistää tekoälyn käyttöä ja kehitystä, mutta samalla varmistaa unionin oikeushyvien toteutuminen.<sup>3</sup>

Euroopan komission Valkoinen kirja tekoälystä toteaa, että digitaalitekniologioiden tullessa yhä keskeisemmäksi osaksi ihmisten elämää, ihmisten tulee voida luottaa teknologiaan. Luottamus on edellytys teknologian käyttöönotolle.<sup>4</sup> Luottamuksen varmistamiseksi eurooppalaisen tekoälyn tulee perustua arvoihimme ja perusoikeuksiimme, kuten ihmisarvoon ja yksityisyyteen.<sup>5</sup> Luottamuksen rakentamiseksi Valkoisessa kirjassa esitellään tekoälyn sääntelykehikseksi “luottamuksen ekosysteemi”, joka perustuu korkean tason asiantuntijaryhmän julkaisemiin luotettavaa tekoälyä koskeviin eettisiin ohjeisiin.<sup>6</sup> Ohjeissa luotettavan tekoälyn perustaksi määritellään neljä perusoikeuksista kumpuavaa eettistä periaatetta, jotka ovat ihmisen itsemääräämisoikeuden kunnioittaminen, vahinkojen välttäminen, oikeudenmukaisuus ja selitettävyyden.<sup>7</sup> Näiden periaatteiden ja siten luotettavan tekoälyn toteuttamiseksi ohjeissa esitellään seitsemän keskeistä vaatimusta, joista yksi on ihmisen toimijuus ja ihmisen suorittama valvonta.<sup>8</sup>

Periaatteen mukaisesti tekoälyjärjestelmien on tuettava ihmisen itsemääräämisoikeutta ja päätöksentekoa. Ihmisen itsemääräämisoikeuden

---

1 Euroopan komissio 2021 & Euroopan parlamentti 2023.

2 Euroopan parlamentti 2019 & Euroopan unionin neuvosto (COM (2021) 206 final).

3 Tekoälysäädös johdanto-osan perustelukappale 6.

4 Euroopan komissio COM(2020) 65 final s. 1.

5 Ibid. s. 2.

6 Ibid. s. 10.

7 Euroopan parlamentti 2019 s. 13-16.

8 Ibid s. 19. Muut vaatimukset ovat 2) tekninen luotettavuus ja turvallisuus, 3) yksityisyyden suoja ja datan hallinta, 4) läpinäkyvyys, 5) monimuotoisuus, syrjimättömyys ja oikeudenmukaisuus, 6) yhteiskunnallinen ja ekologinen hyvinvointi ja 7) vastuuvollisuus.

kunnioittaminen edellyttää järjestelmien mahdollistavan demokraattisen, kukoistavan ja tasa-arvoisen yhteiskunnan toteutumisen tukemalla käyttäjän toimijuutta ja edistämällä perusoikeuksia sekä sallimalla ihmisen suorittaman valvonnan.<sup>9</sup> Ihmisen suorittaman valvonnalla varmistetaan, ettei tekoälyjärjestelmä heikennä ihmisen itsenäistä päätäntävaltaa tai aiheuta muita kielteisiä vaikutuksia.<sup>10</sup>

Tekoälysäädös rakentuu valmisteluasiakirjojen periaatteille ja vaatimuksille.<sup>11</sup> Siten se sisältää huomattavasti ihmisen ja tekoälyjärjestelmän vuorovaikutusta ohjaavia ja määrittäviä säännöksiä. Näin ollen ihmisen ja tekoälyjärjestelmän vuorovaikutuksen voidaan sanoa oikeudellistuvan tekoälysäädöksen myötä.<sup>12</sup> Onnistunut vuorovaikutus on demokratian, oikeusvaltion ja perusoikeuksien toteutumiselle keskeistä ja luotettavan tekoälyn edellytys.<sup>13</sup>

Ihmisen ja järjestelmän vuorovaikutuksen painoarvon vuoksi tavoitteenani tutkielmassa on selvittää, millainen käsitys ihmisen ja järjestelmän vuorovaikutuksesta tekoälysäädöksessä on omaksuttu. Apuna ja vertailukohtana tulen käyttämään ihmisen ja tietokoneen vuorovaikutukseen keskittyneen HCI-tutkimusperinteen (eng. “human-computer interaction research”) mallinnuksia vuorovaikutuksesta. Käsitteen selvittäminen on tarpeen, jotta voin keskustella oikeudellistuneen vuorovaikutuskäsityksen puutteista ja jännitteistä. Tarkastelen tutkielmassa tekoälysäädöstä sisällön analyysin keinoin, ja tutkielmassa hyödynnetään siten empiiristä tutkimusotetta.<sup>14</sup>

## 1.2 Teoreettinen viitekehys, tutkimuskysymykset & metodit

### 1.2.1 Teoreettinen viitekehys

Tutkielman viitekehystenä toimii tiede- ja teknologiatutkimus (eng. “Science and Technology Studies”, STS). Tiede- ja teknologiatutkimus on heterogeeninen ja monialainen suuntaus, joka keskittyy yhteiskunnan, tieteen ja teknologian

---

<sup>9</sup> Ibid. kohta 62.

<sup>10</sup> Euroopan komissio COM(2020) 65 final s. 23.

<sup>11</sup> Tekoälysäädös johdanto-osan perustelukappale 7.

<sup>12</sup> Oikeudellistumisella tarkoitetaan yleisesti yhteiskunnallista kehitystä, jossa säädösten, oikeudenkäyntien ja lakimiesten määrä lisääntyy. Sääntely tihentyy tunkeutumalla uusille elämänalueille, jotka aiemmin ovat olleet vähän säänneltyjä tai sääntelemättömiä.

Oikeudellistumisesta esim. Koivisto 2015 s. 959 & Tarasti 2002 s. 575–585.

<sup>13</sup> Euroopan parlamentti 2019 kohdat 63 & 64.

<sup>14</sup> Ervasti 2017 s. 21.

vuorovaikutuksen tutkimiseen.<sup>15</sup> Yksinkertaistettuna suuntaus keskittyy kysymään, kuinka tiede ja teknologia muokkaavat yhteiskuntaa ja kuinka yhteiskunta muokkaa niitä.<sup>16</sup> Suuntauksen keskiössä on nimenomaan vuorovaikutussuhteen ymmärtäminen.

Keskustelu teknologian vaikutuksista yhteiskuntaan sisältää myös näkökulman, jonka mukaan teknologiat luonnonvoiman tapaan määrittävät yhteiskuntien ja siten myös oikeudellisten käytäntöjen kehityksen. Näkökulmaan sisältyy käsitys vaikutusmahdollisuuksien rajallisuudesta tai olemattomuudesta.<sup>17</sup> Tähän teknologia determinismin (eng. “technology determinism”) teoriaan liittyy vahvasti ajatus, jossa teknologian vaikutukset yhteiskuntaan ja sosiaaliseen todellisuuteen nähdään yksisuuntaisia.<sup>18</sup> Sitoutumalla STS-viitekehykseen hylkään tämän teorian. Suuntauksen viime vuosikymmenten tutkimuksessa onkin omaksuttu teknologian sosiaalisen muotoilun teoria (eng. “social shaping of technology”, SST), jossa korostetaan vaikutusmahdollisuuksia kehitykseen. Teoria huomauttaa, että teknologian kehittäminen ei ole arvovapaata vaan sitä muovaavat ja määrittävät yhteiskunnan hallitsevat sosiaaliset, poliittiset ja taloudelliset arvot. Muutokset yhteiskunnallisessa ilmapiirissä johtavat siten erilaisiin teknologisiin tuloksiin.<sup>19</sup>

Tutkielma on tarkemmin sijoitettavissa osaksi oikeuden ja teknologian tutkimuskenttää.<sup>20</sup> Tutkielman tarkempi positiointi on yhteensopiva STS-viitekehyksen kanssa, sillä ymmärrän oikeuden erityisenä yhteiskunnallisena käytäntönä.<sup>21</sup> STS-viitekehyksestä hyödynnetään eritoten ajatusta vuorovaikutussuhteesta. Tutkielman ankkurointi aiempaan oikeuden ja teknologian tutkimukseen antaa puolestaan tarvittavat käsitteet keskustelulle. Tutkielman tavoitteena on oikeuden, yhteiskunnan ja teknologian kehikossa selvittää etenkin eksplisiittisiä mutta myös implisiittisiä oletuksia, joita tekoälysiisäädös sisältää ihmisen ja järjestelmän vuorovaikutuksesta. Suunnatessani tarkastelun lainsäädännössä omaksuttuihin oletuksiin sitoudun näkökulmaan, jonka mukaan lainsäädännöllä voidaan vaikuttaa teknologissosiaaliseen todellisuuteen. Teknologian ja oikeuden

---

15 Felt et al. 2016 s. 1-2.

16 Law 2016 s. 31.

17 Ks. Winner 1980 s. 121–136.

18 Ibid. s. 121-123.

19 Williams & Edge 1996 s. 865–899.

20 Esim. Hakkarainen et al. 2020 s. 6-7; Cole & Bertenthal 2017 s. 351–371.

21 Evarsti 2017 s. 9-10.

suhde on alituisesti muuttuva ja moniulotteinen, eikä selkeitä syy-seuraussuhteita ole eroteltavissa. Tutkimuskirjallisuudessa tätä suhdetta on kuvattu dynaamiseksi vastavuoroisuudeksi (eng. “dynamic reciprocity”).<sup>22</sup> Oikeudella on mahdollista vaikuttaa merkittävästi teknologiseen kehitykseen, mutta samanaikaisesti se ei ole immuuni vastavaikutuksille.

### 1.2.2 Tutkimuskysymykset ja tutkielman rajaukset

Tutkielmassa keskityn tarkastelemaan ihmisen ja järjestelmän vuorovaikutusta oikeudellistuvana ilmiönä. Tutkimuskysymykseni ovat seuraavat: 1) millainen tekoälysäädöksessä omaksuttu oikeudellistuva vuorovaikutuskäsitys on; 2) miten säädöksessä oikeudellistuva vuorovaikutuskäsitys vertautuu HCI-tutkimuksen vuorovaikutuskäsitykseen; ja 3) miten oikeusjärjestelmän syvärakenne määrittää vuorovaikutuskäsitystä ja kuinka se näkyy teknologiasäätelylle tyypillisinä ongelmina tekoälysäädöksessä?<sup>23</sup>

Tekoälysäädöksessä omaksutun käsityksen selvittämiseksi säädöstä analysoidaan käyttäen apuna HCI-tutkimuksessa omaksuttuja vuorovaikutusmallinnuksia ja käsitteitä (TK1). Kuvattu analyysitapa mahdollistaa hahmotustapojen erojen ja yhtäläisyyksien vertaamisen. Yhtäläisyyksien ja eroavaisuuksien havaitsemisella pyritään yhtäältä arvioimaan, onko lainsäädäntöön omaksuttu toisen tieteenalan tutkimustietoon nojaava perusteellinen käsitys vuorovaikutuksesta vai onko oikeustieteellinen lähestymistapa vuorovaikutukseen erilainen ja missä määrin (TK2). Toisin sanoen, kuinka paljon vuorovaikutuskäsitys muuttuu siirryttäessä HCI-tutkimuksesta lainsäädäntöön. Tutkimuskysymysten 1 ja 2 vastausten valossa keskustellaan, mitkä oikeudelliset rakennelmat ja käsitteet pakottavat tiettyä käsitystä vuorovaikutuksesta oikeudelliseen kontekstiin siirryttäessä (TK3). Pyrin vastaamaan tutkimuskysymyksiin osana suurempaa viitekehystä, jonka mukaan teknologia ja oikeus muuttavat toisiaan vastavuoroisesti ja vastavuoroisuuden vuoksi teknologian säätely on haasteellista.

HCI-tutkimuksen osalta rajaan tarkastelun kahteen vastakkaiseen vuorovaikutusmallinnukseen. Ihmisen ja koneen eriäville kyvyille ja tarkalle

---

<sup>22</sup> Koulu 2020 s. 13.

<sup>23</sup> Myöhemmin käytän seuraavia lyhenteitä: Tutkimuskysymys yksi on “TK1” ja niin edelleen.

tehtävänjaolle perustuvaan “HABA-MABA”<sup>24</sup>-malliin sekä ihmisen ja koneen molempia hyödyttävän yhteistyön “Human-Computer Symbiosis”-malliin. Teen vertailun käyttäen vain näitä kahta mallia, mutta esittelen lisäksi automaation, autonomian ja toimijuuden käsitteiden sisällön HCI-tutkimuksessa, joita käytän apuna sekä vuorovaikutuskäsityksien vertailussa että yleisesti oikeudellisen käsityksen tarkoituksenmukaisuuden arvioinnissa.

Tekoälysäädöksen vaatimuksista tutkielmassa käsittelen erityisesti ihmistoimijuuden ja ihmisvalvonnan velvoitetta oikeudellistuvan vuorovaikutuskäsityksen ilmentäjänä, jota vertaan HCI-mallinnuksiin. Muita vaatimuksia ovat ihmisen toimijuuden ja ihmisen suorittaman valvonnan lisäksi tekninen luotettavuus ja turvallisuus,<sup>25</sup> yksityisyyden suoja ja datan hallinta,<sup>26</sup> läpinäkyvyys,<sup>27</sup> monimuotoisuus, syrjimättömyys ja oikeudenmukaisuus,<sup>28</sup> yhteiskunnallinen ja ekologinen hyvinvointi<sup>29</sup> sekä vastuuvollisuus.<sup>30</sup>

Vaikka on selvää, että vuorovaikutuskäsitykseen vaikuttaa jossain määrin kaikki säädöksen neljä periaatetta ja seitsemän vaatimusta, keskityn tutkielmassa käsittelemään ihmistoimijuuden ja ihmisen valvonnan vaatimusta, koska se redusoi parhaiten aineiston analyysin tulokset ja vuorovaikutuskäsityksen ytimen. Tutkielman oikeudellisessa keskustelussa käytän joitain syrjimättömyyden, monimuotoisuuden ja oikeudenmukaisuuden vaatimusta ilmentäviä esimerkkejä, joiden käyttöä perustelen esimerkkien vahvalla kytkeytymisellä ihmistoimijuuteen.

Vaatimusten lisäksi tekoälysäädöksen valmisteluasiakirjoista ilmenevä karkea riskien jaottelu toimii tässä työssä havainnollistavana rajauksena.<sup>31</sup> Tekoälyn aiheuttamat riskit jaetaan oikeudellisen systematiikan mukaisesti aineettomiin ja aineellisiin vahinkoihin. Esittelen työssäni riskejä, jotka voivat aiheuttaa aineettomia vahinkoja

---

24 HABA-MABA (eng. “Humans are better at, machines are better at”) on eriäviä ominaisuuksia ja vahvuuksia korostava malli.

25 Euroopan parlamentti 2019 s. 17. Sisältää vastustuskyvyn hyökkäyksiä vastaan ja varmuuden, varasuunnitelman ja yleisen turvallisuuden, tarkkuuden, luotettavuuden ja toistettavuuden.

26 Ibid. Sisältää yksityisyyden kunnioittamisen, tietojen laadun ja eheyden sekä tietojen saannin.

27 Ibid. Sisältää jäljitettävyyden, selitettävyyden ja tiedotuksen.

28 Ibid. Sisältää epäoikeudenmukaisten vääristymien välttämisen, esteettömyyden ja kaikkia palvelevan suunnittelun sekä sidosryhmien osallistumisen.

29 Ibid. Sisältää kestävyuden ja ympäristöystävällisyyden, sosiaaliset vaikutukset, yhteiskunnan ja demokratian.

30 Ibid. Sisältää tarkastettavuuden, kielteisten vaikutusten minimoinnin ja niistä ilmoittamisen, valinnat ja muutoksenhaun.

31 Euroopan komissio COM(2020) 65 final s. 11-14.

perusoikeuksille. Sen sijaan en keskity riskeihin, jotka kohdistuvat terveyteen ja turvallisuuteen ja voivat johtaa aineellisiin vahinkoihin ja joita lähestytään erityisesti tuotevastuun ja kuluttajan suojan kautta. Lisäksi kysymykset koskien vastuunpaikantamista tekoälyjärjestelmien arvoketjuissa ovat kokonaan tutkielman viitekehyksen ulkopuolella.

Aineiston esittelyä koskevassa kohdassa asetan tekoälysäädöksen osaksi unionin lainsäädäntökehystä, mutta tutkielmassa ei varsinaisesti keskitytä käsittelemään yhteyttä tekoälysäädöksen vaatimusten ja valitun kontekstin välillä. Toisin sanoen tutkielmassa ei arvioida, miltä osin valittu keinovalikoima vaatimuksiin vastaamiseksi olisi erilainen, jos tekoälysäädöstä ei olisi paikannettu osaksi perusoikeus-, tietosuojaja-, tuotevastuu- ja kuluttajansuojalainsäädäntöä. Tutkielmassa en myöskään esitele vaihtoehtoisia sääntelymalleja tai pyri tekemään de lege ferenda-suosituksia.

Hypoteesina tässä tutkielmassa on, että tekoälysäädöksen tavoitteet ja tavoitteiden toteuttamiseksi valittu keinovalikoima ohjaavat merkittävästi oikeudellista vuorovaikutuskäsitystä (TK1). Tavoitteet ja keinot ovat puolestaan jo todetulla tavalla pitkälti politiikkadokumenttien mukaisia. Oletettavasti vuorovaikutuskäsityksissä sekä eroja että joitain yhtäläisyyksiä HCI-tradition vuorovaikutusmallien kanssa (TK2). Tutkimuskysymyksellä ja aineistolla ei kuitenkaan kyetä vastaamaan siihen, ovatko mallinnukset omaksuttu tietoisesti toiseen tieteelliseen traditioon perustuen, vai onko yhtäläisyydet sattumalta omaksuttuja. Eroavaisuuksien osalta oletuksena on, että oikeudellinen konteksti tuo muutoksia vuorovaikutuskäsitykseen (TK3).<sup>32</sup>

### 1.2.3 Metodit

Tutkielmassa käytettävät metodit on syytä eritellä tutkimuskysymyksen kohtaisesti, sillä sitoutuminen poissulkevasti yhteen menetelmään ei palvele tutkielman tarkoitusta. Tutkielma on siten metodisesti avoin ja yhdistän erilaisia tutkimuksellisia lähestymistapoja.<sup>33</sup> Ensimmäisen tutkimuskysymyksen osalta lähtökohtana on lainopillinen tutkimusmetodi. Lainopin tiedonintressin mukaisesti tutkimuksen tarkoituksena on selvittää lainsäädännön sisältö oikeudellisen vuorovaikutuksen

---

<sup>32</sup> Nähdäkseni tässäkin yhteydessä en pysty osoittamaan, perustuvatko erot tietoiselle harkinnalle. Viittaan tietoisuudella tässä siihen ymmärretäänkö, mitkä oikeusjärjestelmämme rakenteet aiheuttavat erot vuorovaikutuskäsityksessä oikeudellisessa kontekstissa.

<sup>33</sup> Määttä 2015 s. 2.



käsitystavan osalta.<sup>34</sup> Lähestymistapa on de lege lata. Tiedostan tältä osin, että tekoällysäädös ei ole vielä kirjoitushetkellä voimassa olevaa oikeutta. En kuitenkaan näe, että tämä muodostaa ongelman metodin käytölle, sillä säädös on hyväksytty säännönmukaisessa lainsäädäntömenettelyssä.

Tutkielmassa en selvitä vuorovaikutuksen käsitystapaa lainopillisen tutkimuksen tyypillisten menetelmien eli käsiteanalyysin tai oikeudellisen argumentaation keinoin.<sup>35</sup> Ervasti on esittänyt, ettei positiivisoikeudellisten oppiaineiden piirissä ole estettä yhdistää empiiristä lähestymistapaa lainopin menetelmiin.<sup>36</sup> Siten käytän empiiristä tutkimusotetta välineenä, joka mahdollistaa lainopissa perinteisesti käytettäviä oikeudellisia tulkintamenetelmiä syvällisemmän käsityksen muodostamisen aineistosta.<sup>37</sup> Empirian keinona on laadullinen sisällön analyysi, joka laajasti ymmärrettynä kattaa kaiken laadullisen tutkimusaineiston tiivistämisen ja kategorisoinnin.<sup>38</sup> Sen lisäksi, että empiirinen tutkimusote palvelee välineenä tutkimuskysymyksiin yksi ja kolme vastaamista, se muodostaa itsessään metodin tutkimuskysymykseen kaksi vastaamiselle. HCI mallien ja käsitteiden avulla pystyn tehdä teoriasuuntautuneen sisällön analyysin, joka mahdollistaa samalla vertailun suorittamisen.<sup>39</sup>

Tutkielman kolmannen tutkimuskysymyksen tiedonintressi on sääntelyteoreettinen. Sääntelyteorian tarkoituksena on palvella yhteiskunnallisia tiedonintressejä kysymyksenasetteluilla, joiden avulla on mahdollista arvioida yksittäisten säädösten tai oikeusjärjestelmän tasokkuutta.<sup>40</sup> Sääntelyteoreettinen lähestymistapa mahdollistaa sääntelyn tarkoituksenmukaisuuden kriittisen arvioinnin. Kriittinen arviointi voi pitää sisällään sekä de lege lata että de lege ferenda-tutkimuksen elementtejä.<sup>41</sup> Sääntelyteoreettinen lähestymistapa hyötyy sekä lainopillisesta että empiirisestä tutkimusotteesta, sillä näiden avulla tuotetaan itsetietoisuutta vuorovaikutuksen käsitteellistämistävasta oikeudellisessa kontekstissa. Empirian

---

34 Aarnio 2011 s. 13.

35 Ibid.

36 Ervasti 1998 s. 372 & 2013 s. 21.

37 Määttä 2015 s. 141.

38 Tuomi & Sarajärvi 2003 s. 6.

39 Ibid. s. 116.

40 Määttä 2002 s. 132-133.

41 Ibid.

avulla tapahtuva vertaaminen HCI-tutkimukseen auttaa lainsäädännön laadun reflektoinnissa.

Olen edellä tunnistanut tutkielmalleni teoreettisen viitekehyksen, joten tarkempi metodologinen positiointi ei ole enää tarpeen.<sup>42</sup> Teknologialiitännäiselle empiriselle oikeustutkimukselle ominaiseen tapaan tutkielman kysymyksenasettelut eivät ole asetettavissa perinteiseen oikeudenalajaotteluun.<sup>43</sup> Vuorovaikutuksen laadusta käytävässä oikeudellisessa keskustelussa käytän apuna oikeusjärjestelmälle yhteisiä käsitteitä.

Koulu on esittänyt oikeuden digitalisaation tutkimukselle lupauksen ja tavoitteen. Lupaus on oikeuden itseymmärryksen lisäämisen kautta pystyä vastaamaan käytännön oikeudellisiin haasteisiin, joita oikeuden digitalisaatio tuo. Tavoitteena on sen sijaan oikeudelle asetettujen perustavanlaatuisten vaatimusten täyttäminen muutoksesta huolimatta.<sup>44</sup> Olen omaksunut lupauksen itseymmärryksen lisäämisestä tutkielman tavoitteeksi. Vaikka tutkielmassa toistamani havainnot oikeuden antroposentrismistä ja digitalisaation haasteista oikeuden rakenteille eivät ole uusia, teen näkyväksi ristiriidan vuorovaikutuksen käsitystavassa suhteessa HCI-tutkimukseen. Samalla ihminen-järjestelmä vuorovaikutukseen keskittyvästä tutkimuksesta saadaan tarkemmat käsitteet vuorovaikutuksen kuvaamiselle, joka auttaa luomaan tarkemman kuvan oikeudellistuvasta vuorovaikutuskäsityksestä. Mahdollisesti käsitteistä on hyötyä tulevassa oikeudellisessa keskustelussa.

### **1.3 Aineiston kuvailu & tutkielman rakenne**

#### **1.3.1 Aineisto**

Tutkielmassa keskityn nimenomaan tekoälysäädökseen osana EU:n niin kutsuttua big five- sääntelyä, jonka tarkoitus on luoda avoimet sisämarkkinat datatalouden kasvupotentiaalin hyödyntämiseksi turvallisesti.<sup>45</sup> Tekoälysäädös valikoitui aineistoksi sen ollessa kirjoitushetkellä tuorein ja puhutuin EU-oikeudellinen säädös, joka sisältää sääntelyä ihmisen ja järjestelmän vuorovaikutuksesta. Tekoälysäädös on

---

<sup>42</sup> Vrt. Hirvonen 2011 s. 9-10; Tuomi & Sarajärvi 2003 s. 11.

<sup>43</sup> Koulu 2018 s. 842.

<sup>44</sup> Ibid s. 843.

<sup>45</sup> Euroopan komissio COM(2020) 66 final. Eurooppa haluaa maailmanlaajuisen johtoaseman datavetoisen talouden ja sen sovellusten innovoinnissa. Tekoälysäädös on olennainen osa EU:n datatalouden kasvupotentiaalin hyödyntämistä. Euroopan komissio COM(2020) 65 final s. 2.

laatuun ensimmäinen maailmassa, ja sen odotetaan vaikuttavan merkittävästi alan kehitykseen.<sup>46</sup> Merkittävyyttä lisää Bryssel-efekti, jolla viitataan EU:n sisämarkkinoille laatimien lainsäädäntötoimien vaikuttavuuteen maailman laajuisesti.<sup>47</sup> Tekoälysäädöksen taustalla olevasta virallisaineistosta on luettavissa tällaisen vaikuttavuuden tavoittelusta.<sup>48</sup> Perustelen valintaa myös ihmisen ja tekoälyjärjestelmän vuorovaikutukselle annetulla painoarvolla, joka säädöksen voimaan tulon myötä oikeudellistuu. Onnistuneen vuorovaikutuksen katsotaan olevan viimesijainen keino varmistaa unionin oikeushyvien toteutuminen ja se on luotettavan tekoälyn edellytys.<sup>49</sup>

Tekoälysäädöksen valmisteluprosessi oli vivahteikas ja monivaiheinen. Euroopan komission 2021 keväällä tekemästä ehdotuksesta päästiin alustavaan sopimukseen kolmikantaneuvotteluissa joulukuussa 2023.<sup>50</sup> Tekoälysäädöksen lopullinen teksti hyväksyttiin sekä parlamentin valmistelemissa valiokunnissa sekä neuvoston neuvottelijoiden toimesta helmikuussa 2024.<sup>51</sup> Euroopan parlamentti hyväksyi säädöksen täysistunnossaan maaliskuussa 2024.<sup>52</sup> Tutkielmassa on siten analysoitu lopullisen tekoälysäädöksen epävirallista versiota, sillä virallinen versio on saatavilla vasta juristilingvistien kielihuollon jälkeen.<sup>53</sup> Tutkielman tulosten kannalta epävirallisen version käyttö ei ole ongelmallista, sillä sisällöllisiä muutoksia versioiden välillä ei ole.<sup>54</sup>

### 1.3.2 Tekoälysäädöksen tausta – sääntelyn tarve

Tekoälysäädöksen valmisteluasiakirjoissa perustellaan, miksi unionin alueella on tarve erilliselle teknologiasääntelylle. Perustelua taustoitetaan tasapainotettavien tavoitteiden kautta. Näitä ovat tekoälyteknologian käyttöönoton lisääminen ja sovellutusten tuomat mahdollisuudet sekä toisaalta kuluttajien ja yritysten

---

46 Esim. Euroopan komissio 2023.

47 Bryssel-efektistä on puhuttu erityisesti yleisen tietosuoja-asetuksen (GDPR) vaikuttavuuden yhteydessä. Vaikutus toteutuu pääasiassa suurten kansainvälisten yritysten kautta, jotka adaptoivat toimintansa eurooppalaiseen markkinaan sopivaksi. Ks. Bradford 2019.

48 Euroopan komissio COM(2020) 66 final s. 8-9.

49 Euroopan parlamentti 2019 kohdat 63 & 64.

50 Euroopan parlamentti 2024.

51 Ibid.

52 Euroopan parlamentti 2024 (TA(2024)0138) täysistunnossa (13.3.2024) hyväksytty tekoälysäädös.

53 Ibid. alaviite \*.

54 Kielellisesti niitä on nähdäkseni odotettavissa hyvin vähän. Lisäksi analyysissä ei tarkoituksella kiinnitetty huomiota kohtiin, joissa selkeästi oli kielenhuollollinen virhe. Esimerkiksi, jos komission alkuperäisen version käyttäjän määritelmää käytettiin neuvotteluissa omaksutun käyttöönottajän sijaan.

luottamuksen sekä ihmisten terveyden, turvallisuuden ja perusoikeuksien suojele.<sup>55</sup> Valkoisessa kirjassa mainitaan tekoälyteknologian tuovan uusia mahdollisuuksia esimerkiksi ympäristönmuutoksen hillitsemiseksi. Mahdollisuudet ovat siten yleisellä edulla perusteltavissa. Tekoälyteknologian aiheuttamat riskit yksilölle voivat olla esimerkiksi tekoälyjärjestelmien rikollinen käyttö, läpinäkymätön päätöksenteko, syrjivät algoritmit ja tunkeutuminen yksityiselämään.<sup>56</sup> Riskeistä johtuva luottamuksen puute hidastaa teknologian käyttöönottoa. Säätelykehys lisää kuluttajien ja yritysten luottamusta tekoälyyn. Siten säätelyä tarvitaan, jotta tekoälyteknologian käyttöönotto lisääntyy.<sup>57</sup> Lisäksi EU:n laajuinen säätelykehys tarvitaan, jotta sisämarkkinoiden pirstaloituminen ei vaaranna luottamusta, oikeusvarmuutta ja markkinoille pääsyä, ja siten vaikeuta teknologioiden käyttöönottoa.<sup>58</sup>

Ongelman määrittelyä koskevissa kohdissa riskit luokitellaan aineellisiksi vahingoiksi, jotka kohdistuvat terveyteen ja turvallisuuteen tai aineettomiksi vahingoiksi, jotka kohdistuvat perusoikeuksiin.<sup>59</sup> Riskit perusoikeuksille muodostuvat teknologioiden erityispiirteistä, kuten läpinäkymättömyydestä,<sup>60</sup> monimutkaisuudesta, ennakoimattomuudesta ja osittain autonomisesta käyttäytymisestä. Nämä erityispiirteet voivat vaikeuttaa voimassa olevan lainsäädännön noudattamista ja haitata tehokasta täytäntöönpanoa.<sup>61</sup> Lisäksi tekoälyyn yhdistettynä syrjivällä ennakkoluulolla voi olla suurempi vaikutus verrattuna samanlaiseen inhimillisessä päätöksenteossa esiintyvään ennakkoluuloon. Siksi tarvitaan ihmisen käyttäytymistä säänteleviä sosiaalisen valvonnan mekanismeja.<sup>62</sup>

Riskit turvallisuudelle muodostuvat selkeiden turvallisuussäännösten puuttumisesta. Asianomaisille henkilöille tästä koituvan riskin lisäksi turvallisuussäännösten puute aiheuttaa oikeudellista epävarmuutta yrityksille. Markkina- ja

---

<sup>55</sup> Euroopan komissio COM(2020) 65 final s. 10.

<sup>56</sup> Ibid. s. 1-3.

<sup>57</sup> Ibid. s. 10-11.

<sup>58</sup> Ibid. s. 11.

<sup>59</sup> Terveydelle ja turvallisuudelle aiheutuvista aineellisista vahingoista esimerkkinä on omaisuudelle tai ihmishengelle aiheutuvat vahingot. Perusoikeuksille aiheutuvista aineettomista vahingoista esimerkkinä on syrjintä tai yksityisyyden menettäminen. Ibid. s. 11

<sup>60</sup> Niin sanottu ”musta laatikko” -vaikutus.

<sup>61</sup> Esimerkiksi lainvalvontaviranomaisilla ja asianomaisilla henkilöillä ei ehkä ole keinoja varmentaa, kuinka tietty tekoälyn avustuksella tehty päätös on tehty ja onko asiaa koskevia sääntöjä noudatettu.

<sup>62</sup> Euroopan komissio COM(2020) 65 final s. 10-11.

lainvalvontaviranomaisten näkökulmasta epävarmaa on, miltä osin ne voivat puuttua tekoälytuotteiden turvallisuuteen. Haasteeksi muodostuvat epäselvät toimivaltasuhteet ja puutteelliset valmiudet järjestelmien tarkastamiseen.<sup>63</sup> Samat tekoälyn erikoispiirteet, jotka muodostavat riskit perusoikeuksille, koskevat yhtä lailla vastuuseen ja turvallisuuteen liittyviä kysymyksiä.<sup>64</sup>

### 1.3.3 Tekoälysäädös EU-lainsäädännön systematiikassa

Valmisteluasiakirjoissa todettiin, että EU-lainsäädäntö on kattava tuotevastuun, kuluttajansuojan ja syrjinnän kieltävän lainsäädännön osalta.<sup>65</sup> Lisäksi unionin tietosuojalainsäädäntö koskee neutraalisti kaikkia käsittelyteknologioita ja alakohtaisen erityislainsäädännön määrä on huomattava.<sup>66</sup> Voimassa olevan EU:n lainsäädännön todetaan soveltuvan täysimääräisesti riippumatta käytetystä teknologiasta. Valvonnan tehottomuuden ja riskien mitikoimiseksi sääntelyä nähtiin kuitenkin tarpeelliseksi mukauttaa ottaen huomioon tekoälyn erityispiirteet.<sup>67</sup>

Tekoälysäädös tulee siten sijoittumaan osaksi kuluttajansuojaa, tuoteturvallisuutta ja -vastuuta sekä tietosuojaa, yksityisyyttä ja syrjimättömyyttä koskevaa lainsäädäntöä, mutta ei muun jo olemassa olevan yleissääntelyn tapaan teknologianeutraalisti. Valmisteluasiakirjoista ilmenee selkeästi myös tulevan sääntelyn kohderyhmä. Tekoälynsääntely tullaan kohdistamaan tekoälyn kehittäjiin ja käyttönottajiin. Vaikka tunnistettu ja kattava sääntelykehys on todetulla tavalla teknologianeutraali ja sellaisenaan soveltuu jo kehittäjiin sekä käyttönottajiin, nykyisen lainsäädännön soveltamisen ja toimeenpanon haasteet tekoälyn erikoispiirteiden takia synnyttävät tarpeen sääntelylle.<sup>68</sup>

Sääntelykehyksessä havaittiin kuitenkin myös täydennys tarpeita tiettyjen vaatimusten osalta. Valmisteluvaiheessa talouden toimijoilta kerättiin palautetta korkean tason työryhmän ehdottomista vaatimuksista.<sup>69</sup> Palautekierroksen mukaan

---

<sup>63</sup> Ibid. s. 13-14.

<sup>64</sup> Ibid. s. 14.

<sup>65</sup> Lisäksi tuotteiden ja palveluiden esteettömyysvaatimuksia koskeva direktiivi tulee voimaan 2025. Euroopan parlamentin ja neuvoston direktiivi (EU) 2019/882. Julkisten instituutioiden digitaalisten palveluiden osalta savutettavuus on jo voimassa. Euroopan parlamentin ja neuvoston direktiivi (EU) 2016/2102.

<sup>66</sup> Euroopan komissio COM(2020) 65 final s. 11.

<sup>67</sup> Ibid. s. 15.

<sup>68</sup> Ibid. s. 11.

<sup>69</sup> Euroopan parlamentti 2019 s. 16. Vaatimuksia olivat ihmisen toimijuuden ja ihmisen suorittaman valvonnan lisäksi tekninen luotettavuus ja turvallisuus, yksityisyyden suoja ja datan hallinta,

vaatimukset on otettu huomioon nykyisissä oikeus- tai sääntelyjärjestelmissä, mutta läpinäkyvyyttä, jäljitettävyyttä ja ihmisen suorittamaa valvontaa koskevia vaatimuksia ei ole nimenomaisesti käsitelty voimassa olevassa lainsäädännössä monilla talouden aloilla.<sup>70</sup>

#### 1.3.4 Tekoälysäädöksen sisäinen systematiikka

Tekoälysäädöksen lähtökohdaksi asetettiin riskiperusteisuus, joka on suhteellisuusperiaatteen ilmentymä. Lähestymistavalla varmistetaan sääntelytoimien oikeasuhtaisuus.<sup>71</sup> Korkeariskisille tekoälyjärjestelmille asetetuilla vaatimuksilla, kuten ihmisvalvonnalla, turvataan perusoikeuksien, terveyden ja turvallisuuden toteutuminen ja vaatimukset ovat käytössä olevista keinoista vähiten sisämarkkinoita rajoittavia. Tekoälysäädöksen systematiikan muodostava riskiarvio perustuu kumulatiivisuudelle, jossa on määritelty riskialttiit alat ja niillä tapahtuva riskialtis käyttötarkoitus, joka muodostaa korkeanriskin tekoälyjärjestelmän.<sup>72</sup> Riskialttiiksi on määritelty alat, joilla tyypillisesti toteutettujen toimintojen ominaispiirteet huomioon ottaen voidaan olettaa esiintyvän merkittäviä riskejä.<sup>73</sup> Riskialtista käyttötarkoitusta ei ole tarkemmin valmisteluasiakirjoissa määritelty, mutta tekoälysäädöksestä voimme huomata, että riskiä ei ole, kun tekoälyjärjestelmä ei voi vaikuttaa käsillä olevan päätöksen asiasisältöön. Valmisteluasiakirjat antavat kumulatiivisuuden arvioinnista esimerkiksi terveydenhuollon ajanvarauksessa käytettävän tekoälyjärjestelmän. Vaikka ala on riskialtis, käyttötarkoitus ei ole, eikä kumulatiivisuus toteudu.<sup>74</sup>

Kuten todettu sääntelyn tarpeen katsottiin johtuvan tunnistetuista riskeistä, jotka johtuvat tekoälyteknologian erityispiirteistä. Erityispiirteet aiheuttavat sen, että sinänsä kattavan ja valmiiksi soveltuvan lainsäädäntökehikon valvonta, toimeenpano ja tehokkuus ovat kyseenalaisia. Säädös vastaa siten erityisesti tarpeeseen vahvistaa

---

läpinäkyvyys, monimuotoisuus, syrjimättömyys ja oikeudenmukaisuus, yhteiskunnallinen ja ekologinen hyvinvointi sekä vastuuvollisuus.

<sup>70</sup> Euroopan komissio COM(2020) 65 final s. 11.

<sup>71</sup> Ibid. s. 19.

<sup>72</sup> Ibid. s. 19.

<sup>73</sup> Tekoälysäädöksen liite III. Näitä ovat kriittisen infrastruktuurin hallinta ja käsittely, yleissivistävä ja ammatillinen koulutus, työllistäminen, henkilöstöhallinto sekä itsenäisen ammatinharjoittamisen mahdollistaminen olennaisten yksityisten palvelujen ja julkisten palvelujen ja etujen saatavuus ja käyttö, lainvalvonta, muuttoliikkeen hallinta, turvapaikka-asiat ja rajavalvonta sekä oikeudenhoito ja demokraattiset prosessit.

<sup>74</sup> Euroopan komissio COM(2020) 65 final s. 19.

vaatimukset, joilla erityispiirteiden aiheuttamiin riskeihin puututaan ja sen kautta helpotetaan myös vastuun allokointia tekoälyjärjestelmien arvoketjuissa järjestelmäkehittäjistä käyttöönottajiin.

### 1.3.5 Rakenne

Tutkielmassa edetään siten, että ensin esittelen pääpiirteitä HCI-tutkimusperinteestä yleiskuvan antamiseksi. Seuraavaksi esitelen tutkielman tiedonintressin kannalta olennaisimmat HCI-tutkimuksen käyttämät vuorovaikutuksen käsitteellistämisen tavat ja niiden kehityksen. Tämän jälkeen käydään lyhyesti läpi vuorovaikutusmalleihin läheisesti liittyvät automaation, autonomian ja toimijuuden käsitteet. Tiedostan, että oikeustieteen opiskelijan ei ole mahdollista maisterin opinnäytetyössä käyttää toisen tieteenalan käsitteitä ja tutkimusperinnettä suvereenisti.<sup>75</sup> Tarkoituksena HCI-tutkimuksen esittelyssä on aineiston käsitteellistämisen lisäksi auttaa lukijaa ymmärtämään aihetta oikeuden ulkopuolella, jotta oikeuden vuorovaikutukselle antamat ominaispiirteet voidaan paremmin huomata. HCI-jakson jälkeen esitellään havainnot aineistoista ja vertaillaan aineiston osoittamaa oikeudellista vuorovaikutustapaa suhteessa HCI-tutkimuksen vuorovaikutuskäsitykseen (kappale 3). Sen jälkeen oikeudellista vuorovaikutuskäsitystä arvioidaan oikeusjärjestelmän näkökulmasta, käyttäen apuna oikeuden digitalisaatiosta ja teknologissäätelystä käytävää keskustelua (kappale 4). Viimeisenä tarkoituksena on reflektoida lainsäädännössä omaksuttua vuorovaikutuksen hahmotustapaa ja näin ollen lainsäädännön laatua (kappale 5) . Lopuksi esittelen loppupäätelmät.

---

<sup>75</sup> Ks. Bellantyne 2019 s. 367–395.

## 2 HCI-tutkimus

### 2.1 Tutkimusperinteestä ja historiasta

#### 2.1.1 Tutkimusperinteen synty

Ihmisen ja tietokoneen vuorovaikutuksen tutkimus on 1980-luvun alussa muotoutunut tieteenala. HCI-tutkimus on alusta alkaen ollut monitieteistä mutta perinteen “akateeminen koti” on vahvasti tietojenkäsittelytieteissä.<sup>76</sup> HCI-tutkimus on alusta alkaen ammentanut myös kognitiotieteistä ja inhimillisten tekijöiden suunnittelusta.<sup>77</sup> HCI-tutkimuksella on monia muitakin ilmenemismuotoja, mutta sen olemus voidaan pelkistää suunnittelutieteeksi.<sup>78</sup>

Kysynnän tieteenalan syntymiselle loivat 1970-luvun loppupuolella yleistyneet henkilökohtaiset tietokoneet (eng. “personal computing”). Henkilökohtaisilla tietokoneilla viitataan yhden kuluttajaloppukäyttäjän käyttöön suunniteltujen digitaalisten tietokoneiden ilmestymiseen.<sup>79</sup> Kuten myöhemmissä kappaleissa esitän ihmisen ja koneen vuorovaikutuksesta oltiin jo ennen tieteenalan etabloitumista esitetty merkittäviä mallinnuksia.<sup>80</sup> HCI-tutkimuksen alkuperäinen kohde ja yhä keskeinen painopiste on käytettävyys.<sup>81</sup> Käytettävyyden tutkimisessa näkyy HCI-tutkimuksen eri tavalla orientoituneet lähestymistavat. Teknisesti suuntautunut orientaatio korostaa käytettävyydessä vaikuttavuutta, tehokkuutta ja käyttäjän tyytyväisyyttä tavoitteiden saavuttamisessa. Käyttäjäkokeemusta korostava orientaatio kiinnittää nimensä mukaisesti huomiota käyttäjän tunnemaailmaan ja kokemukseen järjestelmän käytöstä. Käyttäjäkokeemusta korostava suuntaus on voimistunut viime aikoina.<sup>82</sup> HCI-tutkimuksessa voidaankin puhua eri aalloista.<sup>83</sup>

---

76 Carroll 2009 s. 28.

77 Ibid. s. 20. Kognitiivisella suunnittelulla viitataan tutkimusmenetelmään, jossa kognitiotieteitä kuten neurologiaa ja psykologiaa käytetään hyväksi teknisiä järjestelmiä käyttäjien kognitiivisten prosessien tukemiseksi. s. 3.

78 Koulu & Hakkarainen 2023 s. 43.

79 Editors of Encyclopedia Britannica 2024.

80 Grudin 2012 s. xxxiv- xxxv.

81 Carroll 2012. s. 7. Carrollin mukaan käytettävyyden käsite pelkistettiin sloganiin “helppo oppia ja helppo käyttää”. Tämä banaali yksinkertaistus määritti pitkään tieteenalan identiteettiä ja intergriteettiä sekä vaikutti teknologian kehitykseen laajemminkin.

82 Esim. Douris 2001 s. 3; Koulu & Westerling 2023 s. 9-10.

83 Bødker 2006 s. 1-8.



### 2.1.2 HCI-tutkimus vuorovaikutuksen tutkimuksen historiana

Yksi tiedetaksonomiasta riippumaton tapa kuvata HCI-tutkimusta on kohteensa kautta. Näin teki esimerkiksi Paul Dourish, jonka mukaan tieteenalassa on kyse ihmisen kykyjen ja koneen tuomien mahdollisuuksien vuorovaikutuksen tutkimisesta. Hän kuvaa, kuinka eri painotuksia omaavalla tieteenalalla yksi tapa tuoda yhteen historiallista kehitystä on tarkastella, mitä ihmisen aistia koneen kanssa vuorovaikuttaminen on milloinkin hyödyntänyt.<sup>84</sup>

Vuorovaikutuksen historiaa on mahdotonta erottaa käyttöliittymien historiasta, koska koneen kanssa vuorovaikuttaminen vaatii käyttöliittymän.<sup>85</sup> Näin ollen käyttöliittymän käsite liittyy olennaisesti vuorovaikutukseen ja on muutenkin HCI-tutkimuksen ydinkäsitteitä. Käyttöliittymä on ihmisen ja koneen vuorovaikutuksen leikkauspiste.<sup>86</sup> Teknologisen kehityksen aste on taas ollut tekijä, joka pitkälti määrittää minkälaisia käyttöliittymiä vuorovaikutuksen pinnaksi on voitu suunnitella ja toteuttaa.<sup>87</sup> Luonnollisesti vuorovaikutuksen historia on siten myös teknologian historiaa.

Dourish on tunnistanut neljä teknologian tason määrittämää aikakautta. Nämä teknologian tasot, jotka vaikuttavat vuorovaikutuksen muotoon, ovat jaettavissa elektroniseen, symboliseen, tekstuaaliseen ja graafiseen vaiheeseen. Elektroninen vaihe on anakronistisesti katsottuna kaukana nykyisestä vuorovaikutuksen käsityksestä.<sup>88</sup> Koneet olivat ammattilaisten käytössä olevia, pääosin "hardwaresta" koostuvia tehtäväkohtaisia prototyyppisiä.<sup>89</sup> Seuraava vaihe käsitti siirtymän numeerisesta koneluettavasta symboliseen muotoon, joka mahdollisti koneen kanssa tekemisissä olevan ihmisen käyttäen moninaisemmin inhimillisiä kykyjä.<sup>90</sup>

Symbolinen esitystapa eteni tekstuaaliseen vaiheeseen, jolloin ihminen pääsi itselleen ominaisella tavalla kommunikoimaan koneen kanssa. Tämä mahdollisti edestakaisen palautteen ihmisen ja koneen välillä, jonka myötä nykyaikaisen käsityksemme kanssa yhteensopiva muoto vuorovaikutuksesta syntyi. Nimenomaan edestakaisesta

---

84 Dourish 2004 s. 5.

85 Ibid. s. 5. Dourish kuvaa käyttöliittymää vuorovaikutuskäsityksen heijastuksena.

86 Esim. Koulu et al. 2022 s. 6.

87 Ks. esim. Rogers et al. 2011.

88 Dourish 2004 s. 5-7.

89 Ibid. Ks. myös Carroll 2009.

90 Dourish 2004 s. 7-9.

palautteesta muodostuva vuoropuhelu on pysynyt vuorovaikutuksen idean keskiössä.<sup>91</sup> Graafiseen vaiheeseen siirtyminen oli kaikista merkittävin muutos. Graafisiin käyttöliittymiin siirtyminen tarkoitti yhä moninaisempia ihmiskognition hyödyntämisen tapoja.<sup>92</sup> Samanaikaisesti vuorovaikutuksesta tuli moniulotteista verrattuna tekstuaaliseen vuorovaikutukseen, joka tapahtui lineaarisena komentovirtana. Graafinen esitystapa lisäsi myös informaatiotiheyttä, sillä ikonit mahdollistivat kommentojen visuaalisen esittämisen tekstin rinnalla.<sup>93</sup> Nähdäkseni tämä siirtymä on ollut vuorovaikutuksen tehokkuuteen olennaisesti vaikuttava tekijä.

Ihmisen ja koneen muodostamat järjestelmät projisoivat ympäröivää maailmaa - ne ovat mallinnuksia todellisuudesta sellaisena kuin järjestelmäkehittäjät ne näkevät. Samaan tapaan HCI-tutkimuksessa keskiössä ovat mallinnukset ihmiskognitiosta- ja toiminnasta. Ihmiskognition ja toiminnan mallinuksilla on filosofinen perusta. Tekniset käytännöt ja niiden toteuttamien nojaavat näihin perustuksiin. Filosofiset perustukset eivät ole yhdentekeviä, vaan niiden ymmärtäminen auttaa hahmottamaan, mihin käytännössä on sitouduttu.<sup>94</sup> Esimerkiksi käsitys vuorovaikutuksesta vaikuttaa ratkaisevalla tavalla toteutustapoihin, mittaamenetelmiin ja mahdollisuuksiin, joilla vuorovaikutusta edellyttäviä järjestelmiä arvioidaan.

## 2.2 Vuorovaikutuksen käsitteellistetyt mallit

### 2.2.1 HABA-MABA (“Humans are better at & Machines are better at”)

Varhaisen ihmisen ja koneen vuorovaikutuksen tutkimuksen käsitteellistämisen keskiössä oli ajatus ihmisen ja koneen eriävistä ominaisuuksista. Eriävien ominaisuuksien ja vahvuuksien vuoksi tehokas automaatio vaati tarkoituksenmukaista tehtävien allokoimista ihmisen ja koneen välillä näiden kykyjä vastaavasti.<sup>95</sup> Oman keskeisen panoksensa funktioallokaatioksi kutsutun tutkimusalan alkusysäykselle antoi Paul Fitts, joka kehitti vuonna 1951 niin kutsutun HABA-MABA-mallin.<sup>96</sup> Malli tunnetaan myös Fittsin listauksena.

---

91 Dourish 2004 s. 5-11.

92 Esimerkiksi avaruudellinen päättely ja kuvio- tai mallitunnistus (eng. “spatial reasoning” ja “pattern recognition”).

93 Dourish 2004 s. 11-14.

94 Dourish 2004 s. viii-ix.

95 Wright et al. 2000 s. 335-336.

96 Bradshaw et al. 2017 s. 283.

Funktioalokaatio on yksi järjestelmäsuunnittelun osa-alue ja sillä viitataan strategioihin, joiden avulla toiminnot ja tehtävät tulisi jakaa ihmisen ja teknologian välillä.<sup>97</sup> Yksinkertaistaen funktioalokaatiota hyödyntävissä järjestelmäsuunnittelun strategioissa lähtökohtana on abstraktio toiminnasta, joka ihminen-kone-järjestelmän tulee suorittaa, jonka jälkeen toiminto jyvitetään ihmisen ja koneen välillä molempien vahvuuksien mukaisesti. Tässä jyvityksessä otetaan huomioon tekninen toteutettavuus sekä toisaalta se, minkä toimintojen voidaan “kohtuullisesti olettaa olevan parempi suorittaa ihmisen toimesta”.<sup>98</sup>

Monet nykyisistäkin funktioalokaation menetelmistä käyttävät jonkinlaista HABA-MABA-listausta apuna määrittäessä työnjakoa ja mahdollisten jaotteluiden hyötyhaitta-arvioita.<sup>99</sup> Fittsin listauksen mukaan ihmiset ovat parempia kognitiota haastavissa tehtävissä. Näihin kuuluu havainnointi, arviointikyky, improvisaatio, perusteleminen ja pitkäkestoinen muisti. Vastaavasti koneen kyvykkyys ylittää ihmisen tehokkuudessa, nopeudessa ja toistoa vaativissa tehtävissä sekä samanaikaisissa toiminnoissa ja lyhytaikaisen muistin käytössä.<sup>100</sup> Yleisellä tasolla ihmisen ominaisuuksien ajateltiin olevan omiaan loistamaan nopeasti muuttuvassa ja epävarmassa ympäristössä, kun taas koneen ennalta määritetyssä ja varmassa toimintaympäristössä. Osin tälle huomiolle perustui ajatus ihmisestä koneen valvojana.<sup>101</sup> Ihmisen rooli koneen valvojana syntyi samanaikaisesti “automatisoidaan kaikki”-näkömyksen (eng. “automate everything”) kanssa. Tämän näkömyksen keskiössä oli, että kaikki voidaan ja tullaan automatisoimaan asteittain. Automaatio nähtiin tapana päästä ihmisen virheiden ikeestä.<sup>102</sup>

Fittsin listaus perustuu tarkkarajaiseen työnjakoon ja tehtävien jyvittämiseen. Tosiasiassa monet tehtävistä vaativat osatehtävien suorittamista, jolloin kokonaisuus on harvemmin jommankumman vuorovaikutukseen osallistuvan käsissä.<sup>103</sup> Kritiikin kohteena on lisäksi ollut ne oletukset, joille malli perustuu sekä puutteellinen abstraktioiden kuvaus, joiden perusteella järjestelmät suunnitellaan. Tutkimuksissa on osoitettu, että hierarkkinen tehtäväkuvaus, standardoidut toimintamenetelmät ja

---

97 Roth et al. 2009 s. 200.

98 Wright et al. 2000 s. 336.

99 Ibid.

100 Fitts 1951 s. 5-12.

101 Laundry 2013 s. 778.

102 Ibid. s. 778-779.

103 Ibid. s. 779.

organisaatiokaaviot ovat harvoin riittäviä kuvaamaan tosiasiallisten tehtävien kompleksisuutta.<sup>104</sup>

Funktioalokaatio oli omiaan luomaan dikotomian ihmisen ja koneen vuorovaikutuksen hahmottamiseen.<sup>105</sup> Tämä dikotomia elää vahvana uusista vuorovaikutuksen käsitteellistämismalleista huolimatta, sillä kritiikistä huolimatta funktioalokaatiota ja eriäville vahvuuksille perustuvaa lähestymistapaa on käytetty paljon käytännön työssä.<sup>106</sup> Osa Fittsin listauksessa esitetyistä vahvuuksista vaikuttaa yhä paikkaansa pitäviltä. On esimerkiksi kiistatonta sanoa koneiden olevan laskentateholtaan ylivoimaisia.<sup>107</sup> Järjestelmien mahdollisuudet päättelyyn ja niiden toimintakyky muuttuvassa ympäristössä ei kuitenkaan yhtä yksiselitteisesti häviä ihmisen vastaaville kyvyille.<sup>108</sup> HABA-MABA listausta arvioitaessa on hyvä muistaa konteksti, jossa malli on luotu. Siihen on vaikuttanut yhtäältä poliittinen ilmapiiri, teknologian kehityksen aste ja tieteenalan menetelmien kehittyneisyys.<sup>109</sup> Funktioalokaatio istuu huonosti yhteen nykyaikaisen ihmisen ja koneen yhteistyötä korostavan lähestymistavan kanssa.

### 2.2.2 Man-Machine Symbiosis

Toisen ihmisen ja koneen vuorovaikutuksen käsitteellistämisen tavan esitteli J.C.R. Licklider 1960-luvulla. Malli oli aikanaan tieteisfiktiiivisenäkin pidetty visionäärinen ajatus ihmisen ja koneen symbioottisesta suhteesta. Lickliderin artikkelin mukaan tulevaisuuden vuorovaikutus ihmisen ja koneen välillä muodostuisi molempia hyödyttävästä läheisestä yhteistyöstä. Hän esitteli symbioosin käsitteen toivoen sen helpottavan ja edesauttavan symbioottisen suhteen mahdollistavien järjestelmien kehittämistä.<sup>110</sup> HCI-tutkimusorientaatioissa on näin ollen ollut hyvin varhaisesta vaiheesta alkaen vaihtoehtoisia tapoja vuorovaikutuksen hahmottamiselle.

Lickliderin mallin mukaan symbioosin tavoitteena olisi antaa koneiden luoda ajattelua samalla tavalla kuin ne luovat ratkaisuja ennalta määritettyihin ongelmiin sekä mahdollistaa ihmisen ja koneen yhteistyö monimutkaisissa ja kompleksissa

---

<sup>104</sup> Wright et al. 2000 s. 336.

<sup>105</sup> Koulu 2020 s. 18-20.

<sup>106</sup> ks. Fuld 2000 s. 217-233 ja Corbridge & Cook 1997 s. 73-88.

<sup>107</sup> Landry 2013 s. 778. Toisaalta kaikkien ominaisuuksien osalta teknologian kehityttyä ei voida tyytyä yksiselitteiseen jakomalliin.

<sup>108</sup> Stanford University 2024 s.5 & 78-82.

<sup>109</sup> Koulu 2020 s. 19-20.

<sup>110</sup> Licklider 1960 s. 4.

tilanteissa reaaliaikaisesti ilman joustamatonta ja ennalta määrättyä mallia.<sup>111</sup> Mallia on tavoiteltu viime vuosikymmenillä, mutta aikaisemmin teknologian riittämättömyys, laskentatehon ja tarkan tieteellisen käsityksen puute symbioosimallin sisällöstä ovat estäneet päämäärän toteutumisen.<sup>112</sup>

Ihmisen ja koneen symbioottiselle suhteelle ei ole HCI-tutkimuksessa vakiintunutta yhtä määritelmää. Niin kuin ei ole symbioottiseen suhteeseen olennaisesti liittyvillä muillakaan vuorovaikutusta kuvaavilla termeillä.<sup>113</sup> Havainnollistavaa on vuorovaikutuksen evoluutiota kuvaavien käsitteiden muutos. Alun perin ihmisen ja koneen vuorovaikutusta kuvattiin yhteiselona (eng. “coexistense”), jossa ihmiset ja koneet jakavat samanaikaisesti työtilan.<sup>114</sup> Kummallakin saattaa olla omat tehtävät, jotka ovat itsenäisiä toisiinsa nähden. Yhteiselosta siirryttiin yhteistyöhön (eng. “co-operation”), jossa yhteisen työtilan lisäksi koneet avustavat ihmistä, jossain näille yhteisessä tavoitteessa.<sup>115</sup> Nykyistä vuorovaikutuksen tasoa voidaan kuvata käsitteellä (eng. “collaboration”). Kaikki tutkijat eivät tee eroa näiden kahden yhteistyön käsitteen välillä, mutta jälkimmäisellä voidaan viitata korkeampaan yhteistyön tasoon, johon liitetään roolien dynaaminen jako.<sup>116</sup>

Ihmisen ja koneen korkean tason yhteistyön etuna on tehokkuus. Ihanteellisesti riittävä vuorovaikutus vähentää yksittäisen osallistujan panosta ja mahdollistaa molempien vahvuuksien hyödyntämisen optimaalisella tavalla.<sup>117</sup> Korkean yhteistyön järjestelmille ominainen roolien dynaaminen jako alleviivaa järjestelmiltä edellytettävää joustavuutta.<sup>118</sup> Viime vuosikymmenen tutkimukset korostavat optimaalisimpien ihminen-kone-järjestelmien (“HCS”)<sup>119</sup> perustuvan dynaamisuudelle ja siten niissä korostuu yhteistyö ilman kummankaan osapuolen valta-asemaa suhteessa toiseen.<sup>120</sup> Parhaimmillaan korkean yhteistyön järjestelmien

---

<sup>111</sup> Ibid. s. 5. Luoda ajattelua tarkoittaa tässä yhteydessä päättelyä.

<sup>112</sup> Hale et al. 2007 s. 1343.

<sup>113</sup> Inga et al. 2023 s. 1-3.

<sup>114</sup> Schmidtler et al. 2015 s. 83-95.

<sup>115</sup> Inga 2023 s. 2.

<sup>116</sup> Ibid.

<sup>117</sup> Curioni et al. 2019 s. 30.

<sup>118</sup> Ibid.

<sup>119</sup> Viimeaikainen tutkimus vilisee erilaisia kirjainlyhennelmiä vanhoista ja nykyaikaistetuista käsitteistä. Joissain tutkimuksissa perinteisemmän “human-computer-system”-termin sijaan on käytetty “sociotechnical system” (STS)-termiä korostamaan nimenomaan järjestelmän sosiaalista ulottuvuutta. Ks. Boy 2023 s. 1.

<sup>120</sup> Inga 2023 s. 3.

kanssa vuorovaikutuksessa ihmisen on mahdollista kokea “me-toimijuutta”.<sup>121</sup> Ihmisen ja koneen symbioottista yhteistyötä korostavat näkemykset ja viimeaikainen tutkimus puoltavat siten vuorovaikutuksen kehitystä hierarkiaa vähentävään suuntaan verrattuna Fittsin listaukseen. Symbioottisen suhteen korostus vastaa autonomisten järjestelmien lisääntymisen trendiä.<sup>122</sup>

## 2.3 Malleihin ja vuorovaikutukseen olennaisesti liittyvät käsitteet

### 2.3.1 Automaatio

Automaation klassisena tavoitteena on korvata ihmisen manuaalinen ohjaus, suunnittelu ja ongelmanratkaisu järjestelmillä.<sup>123</sup> Erityisesti funktioalokaation kannattajien keskuudessa korostui “automatisoidaan kaikki”-filosofia.<sup>124</sup> Ihmisen oli tunnistettu olevan heikoin lenkki ja virheiden aiheuttaja monien järjestelmien käytössä. Ihmisen korvaaminen luotettavalla automaatiolla oli kiehtova ajatus. Tämä oli houkuttelevaa myös kustannustehokkuuden näkökulmasta.<sup>125</sup>

1970-luvun loppupuolelle tultaessa automaation ja ihmisen yhteensovittamisen haasteisiin havahduttiin.<sup>126</sup> Lianne Bainbridge perusti tunnetun kirjoituksensa Automaation ironiat (1983) (eng. “Ironies of Automation”) väitteelle, jonka mukaan insinöörien kiinnostus ihmisfaktoreita kohtaan on osoitus automaation ironiasta, mitä edistyneempi automaatiojärjestelmä on, sitä oleellisempi ihmisoperaattorin kontribuutio voi olla.<sup>127</sup> Pääasiallinen huomio Bainbridgen kirjoituksessa kohdistui kahteen suunnittelijaan liitettävään ironiaan, joiden ympärille kritiikki rakentuu. Suunnittelijan näkökulmasta ihminen on järjestelmissä virheille altistava tekijä. Samanaikaisesti suunnittelijan virheet muodostavat merkittävän virheiden lähteen. Lisäksi suunnittelija, joka haluaa poistaa ihmisen, päätyy jättämään tälle ne tehtävät, joiden automaatiota ei osaa suunnitella. Näin ollen ihmisoperaattorille jää epäkiitollinen tehtävä suorittaa mielivaltaisesti valikoitunut tehtäväkokonaisuus ilman perusteellista harkintaa siitä, kuinka ihmistä tuetaan tehtävissään.<sup>128</sup> Toisin

---

<sup>121</sup> Jenkins et al. 2021 s. 14-15.

<sup>122</sup> Hoc 2000 s. 838-841.

<sup>123</sup> Bainbridge 1983 s. 775.

<sup>124</sup> Funktioalokaatiota on kuvattu tässä mielessä myös jämäperiaatteeksi. Periaate viitataan siihen, että kaikki toiminnot, jotka ekonomisesta ja teknisestä näkökulmasta kannattaa ja on mahdollista automatisoida, tulisi automatisoida. Ks. Wesche & Sonderegger 2019 s. 201.

<sup>125</sup> Landry 2013 s. 779.

<sup>126</sup> Ibid. Aluksi haasteet tunnistettiin ilmailun kontekstissa.

<sup>127</sup> Bainbridge 1983 s. 775.

<sup>128</sup> Ibid.

sanoen haastavimmat, irralliset ja vaikeimmin abstraktoitavat tehtävät jäävät ihmiselle, jolloin on epätodennäköistä, että operaattori voisi toimia hyvin näissä olosuhteissa.

Automatisoiduissa järjestelmissä ihmiselle jää perinteisesti kaksi roolia, valvoa järjestelmää ja tarvittaessa ottaa järjestelmän kontrolli itselleen. Molempiin tehtäviin Bainbridge liittyy osittain päällekkäistä kritiikkiä ihmiskognition mahdollisuuksista tehtävään. Kritiikki kohdistuu nimenomaan ihmisen kykyihin valvoa ja puuttua. Nostan tässä nykykeskustelun kannalta mielestäni olennaisimmat haasteet. Haasteita on esimerkiksi siinä, että kone tekee päätöksiä laskentatehonsa puolesta huomattavasti ihmistä nopeammin, ihmisen on mahdotonta reaaliaikaisesti valvoa, että järjestelmä seuraa sille asetettuja sääntöjä oikein. Yksi ongelma on myös ihmisvalvojan keskittymiskyvyn heikkous. Motivoituneenkin ihmisen on todettu ylläpitävän huonosti keskittymistä kohteeseen, jossa tapahtuu vähän yllättäviä asioita. Osittain tämän takia järjestelmien tulee tyypillisesti varoittaa ihmistä vikatilanteista. Useimmin kuitenkin varoitustavat hämmentävät valvojaa ja näin ollen hankaloittavat järjestelmän toimintaan puuttumista ennemmin kuin auttavat operaattoria siinä. Viimeisenä heikkoutena mainitsen valvojan ihmisen pätevyyden ja ammattitaidon. Ihmisen pitkäkestoinen muisti ei ole aukoton. Ihannetilanteessa, jossa järjestelmän toimintaan tulee puuttua harvoin, ihmisen voi olla vaikea menestyksekkäällä tavalla suoriutua tehtävästä, koska ihmisellä voi olla vain manuaalista saadut ohjeet tukena.<sup>129</sup>

Kuten aiemmin todettiin, ihmisen valvojan kontrollin käsite syntyi samoihin aikoihin "automatisoidaan kaikki"-filosofian kanssa.<sup>130</sup> Se syntyi tarpeesta kontrolloida etänä etenkin kuussa olevia kulkuneuvoja. Tiedonsiirron hitauden vuoksi huomattiin, että tietyt ohjaussilmukat (eng. "control loops") on tehokkaampi hallinnoida automatisoidusti jättäen ihmiselle vain valvojan roolin. Paradigman muutos heijasteli ihmisen muuttunutta työnkuvaa järjestelmiin nähden. Ihmisestä tuli operaattorin sijaan automaatiojärjestelmien johtaja. Tämä paradigma muovasi olennaisesti ihmisen roolia, sillä käsite mahdollisti kontrolliin perustuvan teoriakehityksen synnyn ja sen soveltamisen ihmisen ja koneen vuorovaikutukseen.<sup>131</sup> Huolimatta jo

---

129 Bainbridge 1983 s. 775-77.

130 Landry 2013 s. 779; Ks. myös Sheridan & Ferrell 1975 s. 1-18; Sheridan & Parasuraman 2005 s. 92.

131 Landry 2013, s. 779-780.

esiin tulleista rajoitteista ihmisen ja automaation vuorovaikutuksessa automaatiota ajettiin edelleen.<sup>132</sup> Kautta automaatioon tähtäävän teknologian historian, poliittinen ilmapiiri ja yhteiskunnan suhtautuminen on vaikuttanut teknologian ja ihmisen vuorovaikutuksen muotoutumiseen.<sup>133</sup> Esimerkiksi kylmän sodan aikainen ideologinen taistelu näkyi myös avaruuskilpailussa ja ihmisen osallisuuden säilyttäminen ylläkuvatulla tavalla kontrolloivassa roolissa oli poliittisestikin tärkeää.<sup>134</sup>

### 2.3.2 Autonomia

Teknologinen kehitys on vaatinut automaation käsitteen rinnalle käsitteen autonomiasta. Vaikka äkkiseltään automaation voitaisiin ajatella olevan autonomista siinä mielessä, että se suorittaa tietyn tehtävän osallistui ihminen prosessiin tai ei, automaatio ei riitä kuvaamaan nykYTEknologian kyvykkyyttä. Automaation ajatellaan viittavan järjestelmiin, jotka seuraavat deterministisesti ja kaavamaisesti tiettyjä käskyjä ja toteuttavat suunnitellun ja rajoitetun joukon ennalta määritellyjä tuloksia. Autonomialla puolestaan viitataan järjestelmiin, jotka pystyvät oppimaan, luomaan, kehittymään ja näin ollen muuttamaan toimintaansa toiminnastaan ja toimintaympäristöstään saadun tiedon seurauksena.<sup>135</sup> Käsitteiden ei ajatella olevan toisiaan poissulkevia vaan niiden voidaan nähdä heijastavan eri kehityksen asteita.<sup>136</sup> Automaation ja autonomian käsitteiden välillä olennaista on teknologian vapauden aste sen suorittaessa tehtäviään.<sup>137</sup>

Sosio-tekni- sen eli ihmisen ja koneen muodostaman järjestelmän vuorovaikutuksen suunnittelussa olennaista on toimijuuden jakautuminen. Tämä vaikuttaa järjestelmän hallintoi-ntiin ja vastuunjakautumiseen. Siten ihmisen ja järjestelmän yhteistyön ymmärtäminen ei kosketa vain järjestelmäsuunnittelijoita. Vuorovaikutuksen ominaispiirteiden ymmärtäminen on teknologian sosiaalisten, oikeudellisten ja eettisten vaikutusten takia tärkeää yhä laajemmalle joukolle

---

<sup>132</sup> Ibid.

<sup>133</sup> Ibid. s. 779. Myös Laundryn mukaan huolimatta “automate everything”-filosofiasta poliittinen ilmapiiri ei ollut ikinä kypsä itsestään lentäville lentokoneille.

<sup>134</sup> Koulu 2020 s. 19-20.

<sup>135</sup> Hancock 2017 s. 284.

<sup>136</sup> Ibid. s. 285.

<sup>137</sup> Simmler & Frischknecht 2021 s. 242.



toimijoita.<sup>138</sup> Siksi HCI-tutkimuksessa on koitettu pelkistää tutkimusorientaatiossa esiintyviä autonomian ulottuvuuksia taksonomian avulla.

Autonomian tasojen tunnistamisessa apuna on käytetty neljää eri ulottuvuutta. Läpinäkymättömyys tarkoittaa, onko vuorovaikuttajalla mahdollisuutta seurata järjestelmän päättelyketjua. Toisin sanoen, kuinka syöte (eng. “input”) johtaa tulokseen (eng. “output”). Ennalta määräämättömyys viittaa siihen, johtaako sama syöte aina samaan tulokseen. Adaptaatiolla puolestaan viitataan järjestelmän mahdollisuuteen oppia kokemuksistaan. Avoimuus tarkoittaa kykyä muuttaa syötettä muiden kanssa vuorovaikuttamisen seurauksena.<sup>139</sup> Kaikista kehittyneimpien järjestelmien katsotaan olevan läpinäkymättömiä, ennalta määräämättömiä, adaptiivisia ja avoimia.<sup>140</sup> Tällaiset järjestelmät toimivat korkealla autonomian asteella. Vaikka esitetty taksonomia kuvaa järjestelmien autonomiaa teknisten ominaisuuksien kautta, se pitää sisällään ihmisnäkökulman. Jaottelu tehdään sen mukaan, missä määrin teknologian toimet ovat vuorovaikutuksessa olevalle ihmiselle ymmärrettäviä. Näin ollen teknisen autonomian tasot sisältävät ihmisen näkökulman ihmisen ja järjestelmän välisestä yhteistyöstä.<sup>141</sup>

Nykyaikainen ihminen-kone vuorovaikutus tarkoittaa yhä enenevässä määrin vuorovaikutusta autonomisen järjestelmän kanssa, joka suorittaa tehtäviä ihmisen puolesta.<sup>142</sup> Ihmisen rooli vuorovaikutusta ohjaavana tahona vähenee merkittävästi tai katoaa kokonaan.<sup>143</sup> Vuorovaikutus ihmisen ja autonomisen järjestelmän kesken on olennaisesti erilaista kuin automaattisten järjestelmien. Autonomisten järjestelmien oppimis- ja muuttumiskyvyn vuoksi vuorovaikutuksen yllätyksellisyys ja arvoituksellisuus tulee lisääntymään huomattavalla tavalla.<sup>144</sup> HCI-tutkimuksessa on tunnistettu dynaamisuuden olevan koetun luottamuksen kannalta olennainen seikka.<sup>145</sup>

---

<sup>138</sup> Ibid. s. 239-240.

<sup>139</sup> Ibid. s. 244-245.

<sup>140</sup> Ibid. Muut tasot lähtien automaattisimmasta autonomisempaan ovat: 1) Ennalta määritetty tekninen järjestelmä, joka on nimensä mukaisesti ennalta määritetty, läpinäkyvä, sopeutumaton ja suljettu; 2) Läpinäkymätön järjestelmä, jonka muut piirteet ovat samat kuin tasolla 1; 3) Ennalta määrittelemätön järjestelmä, joka on muutoin läpinäkymätön, sopeutumaton ja suljettu. 4) Adaptoituva järjestelmä, joka on ennalta määrittelemätön, läpinäkymätön ja suljettu.

<sup>141</sup> Ibid. s. 242.

<sup>142</sup> de Visser et al. 2018 s. 1409.

<sup>143</sup> Hancock 2017 s. 284.

<sup>144</sup> Visser et al. 2018 s. 1410.

<sup>145</sup> Ibid.

Autonomisia järjestelmiä kohtaan koettu luottamus onkin keskeinen kysymys järjestelmien suunnittelussa. Autonomisten järjestelmien suunnittelussa tulisi joidenkin näkökulmien mukaan hakea mallia ensisijaisesti onnistuneesta ihminen-ihminen vuorovaikutuksesta ja ottaa ihmisten välisen vuorovaikutuksen standardit suunnittelun lähtökohdaksi. Tulevaisuuden järjestelmäsuunnittelussa saattaakin korostua yhä enemmän yhteiskuntatieteiden merkitys.<sup>146</sup> Tutkijoiden ohella luottamuksen tärkeyttä korostavat kansainväliset järjestöt kuten IEEE.<sup>147</sup> IEEE on julkaissut standardeja ihmiskeskeisyyden, eettisyyden ja luottamuksen varmistamiseksi.<sup>148</sup> Vuonna 2016 järjestö avasi julkisen keskustelun pyytämällä alan toimijoilta kommentteja aloiteohjelmaan *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems*.<sup>149</sup> Ohjelman tavoitteena oli edistää julkista keskustelua siitä, kuinka älykkäät ja itsenäiset teknologiat voidaan mukauttaa moraalisiin arvoihin ja eettisiin periaatteisiin, jotka asettavat etusijalle ihmisten hyvinvoinnin. Toisena tavoitteena oli luoda standardeja ohjelmaan pohjatuen.<sup>150</sup> Ohjelma-aloitteessa korostetaan normien ja arvojen implisiittistä sisään rakentamista järjestelmiin. Aloitteessa otetaan huomion universaaliuden haaste ja tavoitetta korostetaankin normi- ja arvoyhdenmukaisuutena suunnitellun ympäristön kanssa. Suhde ympäristöön liittyy olennaisesti autonomiaan, sillä järjestelmät toimivat osana ympäristöään vaikuttaen itsekin ympäristöönsä. Järjestelmien tulisi myös signaloida toimivansa normien mukaisesti, jotta luottamus voisi rakentua.<sup>151</sup>

### 2.3.3 Toimijuus HCI-tutkimuksessa

Aineiston analyysin ymmärtämiseksi esittelen vielä HCI-tutkimuksen näkemyksen toimijuudesta ja toimijan käsitteestä.<sup>152</sup> HCI-tutkimuksen näkökulmasta toimijuus on ihminen-kone vuorovaikutuksen suunnittelussa olennainen tekijä.

---

<sup>146</sup> Ibid. s. 1410.

<sup>147</sup> Institute of Electrical and Electronics Engineers.

<sup>148</sup> Kaikki projektiin kuuluvat standardit listattuna: IEEE. (2024). *Autonomous Systems Industry Connections*. <https://standards.ieee.org/industry-connections/ec/autonomous-systems/>

<sup>149</sup> IEEE 2016 & IEEE 2018.

<sup>150</sup> IEEE 2016 s.4.

<sup>151</sup> IEEE 2018 s. 33-34. Normien sisään rakentamisen tavoite jaetaan kolmeen osaan: 1) yhteisön normien tunnistaminen; 2) niiden laskennallinen implementointi ja 3) implementoitujen normien vastaavuuden arviointi suhteessa yhteisön normeihin.

<sup>152</sup> Käsite saa erilaisia painotuksia tieteenalasta riippuen, sillä se on suosittu tutkimuskohde niin humanisti- kuin yhteiskuntatieteissäkin. Käsitteellistyksiin vaikuttaa tieteenalojen ontologiset- ja episteemiset lähtökohdat sekä tutkimuskontekstit. Ks. esim. Eteläpelto et al. 2013 s. 45.

Järjestelmäsuunnittelussa yhtenä tärkeimpänä periaatteena on pidetty ihmisen kokemusta kontrollista ja vaikutusmahdollisuuksista vuorovaikutuksessa järjestelmien kanssa.<sup>153</sup> Laajasti ymmärrettynä toimijuudella tarkoitetaan kausaalisuhdetta kahden entiteetin välillä niiden vuorovaikuttaessa keskenään. Tarkemmin kyse on vuorovaikutuksessa tapahtuvien tarkoituksellisten toimintojen suhteesta. Toimijuuteen liitetään perinteisesti kyky tehdä päätöksiä ja toteuttaa niitä. Toimijuus syntyy nimenomaan vuorovaikutuksessa ympäristön kanssa.<sup>154</sup>

HCI-suunnittelu kohdistaa toimijuuden tyypillisesti juuri ihmiseen. Laajempaa määritelmää puoltavissa näkemyksissä korostetaan taas vuorovaikutuksellisen ulottuvuuden tärkeyttä. Vuorovaikutuksellisuuden sisältyä ajatus, että molempien osapuolten, sekä ihmisen että järjestelmän toimintaa ohjaa sisäänrakennettu tavoite. Ajatus toimijuuden ulottamisesta järjestelmälle syntyy siten erityisesti tavoitteen kautta.<sup>155</sup> Toimijuuden käsitteestä on oltu kiinnostuneita erityisesti ohjelmistosuunnittelussa, jossa uutena lähestymistapana on hyödynnetty agenttilähtöistä ohjelmistomallintamista. Niin sanottujen agentti- ja moniagentti järjestelmien suunnittelu on nostanut suosiotaan erityisesti tekoälyteknologioiden parissa. Agentilla tarkoitetaan entiteettiä, joka suorittaa tiettyä toimintoa ympäristössä, josta se on tietoinen ja voi reagoida muutoksiin.<sup>156</sup>

Moniagenttijärjestelmän käsitteellä pyritään tuomaan huomio vuorovaikutuksen merkitykseen. Taustalla on ajatus, että vuorovaikuttajat yhdessä ovat enemmän kuin osiensa summa.<sup>157</sup>

HCI-tutkimuksen parissa vallitsee laaja konsensus siitä, että autonomia on keskeinen osa toimijuutta. Agentilla voidaan tarkoittaa autonomista laitetta, henkilöä tai ohjelmistokomponenttia, joka on vuorovaikutuksessa ympäristönsä kanssa. Joidenkin lähteiden mukaan autonomia määritellään kyvyksi luoda oma agenda ja tavoitella sitä vailla toisen agentin valtaa.<sup>158</sup> Käytännön kannalta autonomiaa ja siten agentin käsitettä ei ole tarpeen arvioida niin tiukasti. Olisi tuskin tarkoituksenmukaista, että ihmisen luomat agentit toimisivat oman agendan mukaan ihmisen tarkoituksen vastaisesti. Pikemminkin agentti tulisi sanan semanttisenkin

---

<sup>153</sup> Gondomar & Mor 2021 s. 138.

<sup>154</sup> Ibid. s. 138-139.

<sup>155</sup> Ibid.

<sup>156</sup> Sterling & Taveter 2009 s. 7.

<sup>157</sup> Ibid. s. 11.

<sup>158</sup> Ibid. s. 7.

merkityksen mukaisesti ymmärtää edustajana. Autonomiaa tuleekin arvioida suhteellisena eikä absoluuttisena.<sup>159</sup> Siksi autonomiaa tärkeämpinä ominaisuuksina agentille voidaan pitää sen paikallisuutta, johon kuuluu kyky havainnoida, vastaanottaa ja lähettää viestejä sekä vaikuttaa ympäristöön. Älykkäällä agentilla näihin kykyihin kuuluu lisäksi reaktiivisuus ja proaktiivisuus tavoitteen saavuttamiseksi.<sup>160</sup>

## 2.4 Jakson loppuksi

Vuorovaikutuksen käsitteellistämistapoja on yllä esiteltyjen HABA-MABA:n ja symbioosin lisäksi paljon muitakin. Jokaisen tarkoituksena on korostaa vuorovaikutuksessa tiettyjä lainalaisuuksia, ominaisuuksia tai piirteitä. Käsitteet vaihtelevat vuorovaikutuksen laadun mukaan esimerkiksi “kone työvälineenä” tai “kone tiimikaverina” (eng. “computers as a tools” tai “computers as a team mates”).<sup>161</sup> Teknologisen kehityksen takia alalla on nähty tarpeelliseksi luoda myös “koneet johtajina”-vuorovaikutuskäsitys.<sup>162</sup> Tarvetta erilaisille mallinnoille perustellaan tieteellisen tiedon ajantasaisuuden ja vastuullisuuden näkökulmalla. Tieteellisten mallinnoksien tulee vastata teknologian kehitystä, jotta tieteen yhteiskunnallinen tehtävä voidaan täyttää. Asianmukaiset käsitteet mahdollistavat yhteiskunnallisen keskustelun. Lisäksi tärkeänä pidetään nykyisten mallien ajantasaisuuden haastamista.<sup>163</sup>

Nähdäkseni HCI-tutkimuksen mallit eivät ole vain deskriptiivisiä kuvauksia vuorovaikutuksesta. Ne määrittävät paljolti lähestymistapoja järjestelmäsuunnitteluun. Tieteenalalle ominaiseen tapaan niiden ympärille luodaan teorioita ja metodologiaa, jotka vastaavasti ohjaavat käytännön suunnittelu- ja kehitystyötä. Tässä työssä ei ole mahdollista, eikä tarkoituksena, tuoda esiin kaikkia malleihin liittyviä problematisointeja tai mallien tutkimiseksi kehitettyjä menetelmäsuuntauksia tai teorioita. Yllä esitellyt vastakkaiset vuorovaikutuksen

---

<sup>159</sup> Ibid. s. 38.

<sup>160</sup> Ibid.

<sup>161</sup> Wesche & Sonderegger 2019 s. 201.

<sup>162</sup> Tarve uudelleenlaiselle vuorovaikutuksen mallintamisen paradigmatte nähdään välttämättömänä, koska nykyinen ja tuleva teknologinen kehitys mahdollistaa tällaiset käänteiset hierarkkiset suhteet. Eli suhteet, joissa ihmiset suorittavat tehtäviä tietokoneiden käskystä. Esimerkiksi Uber-järjestelmä automatisoi useita johtamistoimintoja, kuten tehtävien jaon, vuorosuunnittelun, suorituspalautteen ja palkitsemisen. Ks. *ibid.* s. 198.

<sup>163</sup> Ibid.

mallinnukset heijastavat nähdäkseni sekä HCI-tutkimusalan kehityksen trendejä että teknologian kehitystä. Vain oikeustiedettä opiskelleen kirjoittajan on vaikea tunnistaa, missä määrin teknologian kehitys on johtanut uusiin vuorovaikutusmallinnuksiin toteutustapojen mahdollisuuksien lisääntyessä ja missä määrin tieteellisen keskustelun paradigmanmuutoksen ohjaava vaikutus näkyy käytännön työssä. Kyse on varmaan molemmista ja tyydyn toteamaan painopisteen vuorovaikutuksen mallinnukseen käytettävien viitekehysten osalta siirtyneen symbioottista vuorovaikutusta korostavaan suuntaan.

### 3 Aineiston analyysi

#### 3.1 Kuvaus analyysirungosta- ja menetelmistä

Edellä oleva HCI-tutkimusosio mahdollisti teorialähtöisen sisällönanalyysin.<sup>164</sup> Teoriaosuuden tarkoituksena oli luoda valmiit ja aikaisempaan tietoon perustuvat käsitteellistämiskehykset vuorovaikutuksesta, joita käytettiin analyysin lähtökohtana.<sup>165</sup> Sisällönanalyysia varten tehty runko muodostui siten HCI-tutkimusosiossa esiteltyjä käsitteitä ja niiden osatekijöitä mukaillen. Strukturoitu runko mahdollisti olemassa olevien käsitteellistämismallien, HABA-MABA:n ja symbioosimallin, vertailun uudessa oikeudellisessa kontekstissa.

Alla oleva taulukko 1 kuvaa strukturoitua analyysirunkoa. Analyysi eteni siten, että aineisto koodattiin ensin neljän eri vuorovaikutusmuuttujan mukaan. Näitä olivat roolitus, hierarkia, toimijuus ja tavoitteet. Näiden muuttujien operationalisoinnissa noudatettiin HCI-tutkimuskirjallisuudessa esiin nostettuja havaintoja muuttujien muodostajista.<sup>166</sup> Esimerkiksi roolitukseen liittyvissä huomioissa koodasin aineistosta kohdat, jotka käsittelivät työnjaon pysyvyyttä tai joustavuutta. Hierarkiaan liittyviksi katsoin ne kohdat, jotka liittyivät työnjaon arvottamiseen, valvontaan ja kontrolliin tai indikaattoreihin näiden olemattomuudesta. Toimijuuden osalta kiinnitin huomiota, viittaako mikään toimijuuden myöntämiseen ihmiselle ja järjestelmälle yhdessä, joka HCI-tutkimuksen mukaan olisi mahdollista toiminnan tavoitteen kautta. Tavoitteiksi koodattiin kohtia, jotka käsittelivät vuorovaikutuksen pyrkimyksiä, joko kummallekin omaksi tai yhteiseksi ymmärrettäviä.<sup>167</sup>

Taulukko 1: Strukturoitu analyysirunko

	HABA-MABA	SYMBIOOSI
Roolitus	Staattisuus; Allokaatio; Tarkkarajaisuus	Joustavuus; Reaktiivisuus
Hierarkia	Valvonta; Kontrolli	Ei selvää johtajaa; Sirkulaarisuus

<sup>164</sup> Tuomi & Sarajärvi 2002 s. 116.

<sup>165</sup> Ibid.

<sup>166</sup> Ibid s. 117.

<sup>167</sup> Aineistoa lukiessa ja vuorovaikutusmuuttujia koodattaessa huomioni kiinnittyi merkityksen lisäksi tekoälysäädöksessä käytettyihin sanavalintoihin ja lauserakenteisiin. Sanat kuten työväline ja tukea auttoivat vuorovaikutusmuuttujien jäsentämisessä.

Toimijuus	Ihminen tai järjestelmä erikseen; Toimijuus vain itse luodun tavoitteen kautta	Me-toimijuus; Toimijuus suhteessa ympäristöön (vuorovaikutteinen ja vastavuoroinen, reaktiivinen)
Tavoitteet	Jaotellut & Erilliset; Kohdistuvat osatehtävään	Yhteinen tavoite; “Enemmän kuin osiensa summa”
Lopputulokset	Dikotomia	Tiivis yhteistyö

Aineiston analyysissä käytettiin apuna Atlas ti-ohjelmaa, joka loi automaattisia kuvaajia koodatuista aineiston paloista sekä mahdollisesti koodattujen pätkien lukemisen myös irrallaan kokonaisuudesta. Tämä helpotti yhteenvetojen tekemistä vuorovaikutusmuuttujakohtaisesti. Kaikkienensa havaitsin tekoälysäädöksestä 55 hierarkiaa, 49 roolitusta, 49 toimijuutta ja 21 tavoitteita käsittelevää kohtaa. Osittain kohdat ovat päällekkäisiä, sillä yksittäisistä tekstin pätkistä saattoi olla mahdoton erottaa vuorovaikutusmuuttujia toisistaan.<sup>168</sup>

### 3.2 Dikotomisen vuorovaikutuskäsityksen ilmentyminen

#### 3.2.1 Tarkkarajainen ja staattinen roolitus

Tekoälysäädöksessä ihmisen ja järjestelmän roolijako ilmenee riskiluokituksesta. Säädöksessä arvioidaan, millaisella tehtävien jaolla tekoälyjärjestelmän käyttö tai tekoälyjärjestelmä itsessään muodostaa korkean riskin. Arvioinnissa oleellista on oikeushyvien merkityksellisyys. Kuten alussa tuotiin esiin, säädöksen riskiluokitus muodostuu kumulatiivisista arvioista. Ihmisen ja tekoälyjärjestelmän välinen tarkka ja staattinen työnjako näkyi säädöksen lukuisista yksityiskohtaisista tehtävien kuvauksista.<sup>169</sup> Työnjakoa käsittelevissä kohdissa ei tullut ilmi mahdollisuutta joustavaan tehtävien vaihtamiseen.<sup>170</sup> Tätä ilmentää se, että roolien vaihto tehtävien

<sup>168</sup> Lisäksi luonnollisesti johdanto-osan perustelukappaleissa havaitut kohdat on tiivistetty artiklaosioon, mikä tekee luvuista vain suuntaa antavia.

<sup>169</sup> Esimerkiksi tekoälysäädöksen johdanto-osan perustelukappale 53 & artikla 6(3).

<sup>170</sup> Tutkielman rajauksien mukaisesti tekoälysäädöksen koodaamisessa ei kiinnitetty huomiota ihmisten välisiin roolituksiin, jotka muodostavat keskeisen osan säädöksen tarkoitusta tähtäämällä

suorittamisessa merkitsisi riskiluokituksen nousemista tai kielletyn käytännön muodostumista.<sup>171</sup> Toisin sanoen järjestelmän tehdessä ihmiselle varattuja tehtäviä järjestelmästä tulee korkean riskin tai kielletty järjestelmä. Riskiluokituksen nouseminen muuttaa edelleen ihmisen tehtäviä, lisäämällä esimerkiksi ihmisvalvonnan vaatimuksen.

Lukuisat tehtävänkuvaukset osoittivat myös, että järjestelmän rooli on olla ihmisen työväline ja ihmisen päätöksenteon tukija.<sup>172</sup> Sama havainto järjestelmän roolista valmistelevia ja hallinnollisia tehtäviä tekevänä osapuolena näkyy kielenkäytössä. Säädöksessä puhutaan termeillä kuten “työväline”, “edistävä” ja “palvella” (eng. “tool”, ancillary” ja “serve”).<sup>173</sup> Ihmisen rooli on puolestaan olla tietoinen päätöksentekijä.<sup>174</sup> Roolituksen voidaan nähdä heijastelevan olettamia kyvyistä, joita HCI-tutkimuksen puolella ihmiselle ja järjestelmille varhaisen funktioalokaation oppien mukaan on tunnistettu. Fittsin listauksen mukaan ihmiset ovat parempia kognitiota haastavissa tehtävissä. Vastaavasti järjestelmät ovat parempia toistoa vaativissa ja yksinkertaisissa tehtävissä.<sup>175</sup> Tällaiset säännönmukaisuuksista koostuvat tehtävät muodostavat olennaisimman osan järjestelmälle varatuista tehtävistä. Ne saavat esimerkiksi tunnistaa kaavamaisuuksia ihmisten tekemistä päätöksistä ja raportoida poikkeukset keskiarvossa.<sup>176</sup> Näin järjestelmän avulla voidaan välttää ihmisen kognition virheet mutta säilyttää järjestelmällä sen rooli työvälineenä.

HCI-tutkimuksessa käsitys ihmisen ja järjestelmän kyvyistä on muuttunut 60-luvulla esitetystä listauksesta. Nykyisen HCI-tutkimuksen mukaan tekoälyjärjestelmän ja ihmisen vuorovaikutuksessa kadotetaan potentiaalia staattisella ja joustamattomalla tehtäväalokaatiolla.<sup>177</sup> Tekoälysäädöksen roolitus vaikuttaakin perustuvan olettamille kyvyistä, jotka pätevät pikemmin automaattisiin kuin autonomisiin datavetoisiin järjestelmiin. Kuitenkin tekoälyjärjestelmien kyvykkyys tunnustetaan riskikuvauksissa ja toisaalta tekoälyjärjestelmien toiminnallisuuksille asetetuissa

---

vastuunpaikantamiseen tekoälyjärjestelmien arvoketjuissa ja oikeusvarmuuden lisäämiseen toimijoiden välillä.

<sup>171</sup> Ibid.

<sup>172</sup> Esim. tekoälysäädös johdanto-osan perustelukappale 61.

<sup>173</sup> Esimerkiksi johdanto-osan perustelukappaleet 6, 27, 59 & 61.

<sup>174</sup> Esimerkiksi johdanto-osan perustelukappaleet 20, 72 & 73.

<sup>175</sup> Fitts 1951 s. 5-12.

<sup>176</sup> Esimerkiksi tekoälysäädöksen johdanto-osan perustelukappale 53

<sup>177</sup> Curioni et al. 2019 s. 30.



vaatimuksissa. Järjestelmien on esimerkiksi kyettävä toimimaan reaktiivisesti monimutkaisessakin ympäristössä ja huolimatta ihmisen virheellisestä tai epäjohtonmukaisesta toiminnasta vuorovaikutuksessa järjestelmän kanssa.<sup>178</sup> Järjestelmät siis kykenevät paljon muihinkin kuin yksinkertaisiin ja toistoa vaativiin tehtäviin. Tekoälysäädöksessä olisi siten ollut mahdollisuus omaksua vuorovaikutuskäsitys, jossa korostetaan HCI-tutkimuksen suosimaa allokaation tarkoituksenmukaista avoimuutta vanhan dikotomisen käsityksen sijaan. Oikeudellistuva vuorovaikutuskäsitys ei kuitenkaan vastaa tässä HCI-tutkimuksen nykykäsitystä, sillä säädöksen politiikkadokumentteissa valittu lähestymistapa luotettavuuden takaamiseksi edellyttää, että järjestelmien roolin on pelkistytävä ihmisen apuvälineiksi ja ihmisen päätöksenteon tukijoiksi.

### 3.2.2 Ihmisvalvonnan vaatimus luo jyrkän hierarkkian

Ihmisen toimijuuden (eng. “human agency”) ja ihmisvalvonnan (eng. “human oversight”) vaatimuksen ollessa yksi seitsemästä vaatimuksesta luotettavalle tekoälylle on eksplisiittisesti selvää, että säädöksessä korostetaan hierarkiaa järjestelmien ja ihmisen vuorovaikutuksessa. Ihmisvalvonnan vaatimuksesta on tekoälysäädöksessä oma artikla 14 ja sen lisäksi ajatus ihmisen kontrollista järjestelmään leikkaa koko säädöksen läpi. Se on nähdäkseni yksittäisistä vuorovaikutusmuuttujista kaikkein eniten vuorovaikutuksen laatua määrittävä tekijä. Ihmisvalvonnan vaatimuksella tarkoitetaan yksinkertaisesti sitä, että ihmisen täytyy asianmukaisesti pystyä kontrolloimaan ja valvomaan tekoälyjärjestelmiä.<sup>179</sup>

Kuten todettu HCI-tutkimuksen paradigmanmuutos ihmisestä järjestelmien operaattorista järjestelmien valvojaksi syntyi funktioallokaation ja “automasoidaan kaikki”-filosofian yhteydessä. Kuussa olevien kulkuneuvojen ohjaussilmukat olivat tehokkaampi automatisoida tiedonsiirron hitauden vuoksi ja ihmiselle jäi vain valvojan rooli.<sup>180</sup> “Automasoidaan kaikki” lähestymistavassa pyrittiin muutenkin vapauttamaan ihminen järjestelmän operaattorin roolista ja kun ihmiselle ei ongelmitta keksitty vuorovaikutuksessa korvaavaa roolia, ihminen asetettiin järjestelmien valvojaksi. Ihmisestä ei oltu poliittisenkaan ilmapiirin vuoksi valmiita

---

<sup>178</sup> Itseasiassa tämä toimintavarmuus on yksi seitsemästä korkean riskin järjestelmille asetetuista vaatimuksista. Ks. esim. johdanto-osan perustelukappale 75.

<sup>179</sup> Tekoälysäädös artikla 14.

<sup>180</sup> Esim. Laundry 2013 s. 779.

luopumaan.<sup>181</sup> Näin ollen nähdäkseni voidaan sanoa, että ihmisvalvonnan käsite on alusta asti ollut jossain määrin puutteellisesti muotoiltu ja sen synty on ollut poliittisten motiivien värittävä myös HCI-tutkimuksen puolella.

Vaikuttaa siltä, että nyt oikeudellistuvassa vuorovaikutuskäsityksessä ihmisvalvonta omaksutaan osittain samasta syystä. Tehokkuuden ja tekoälyjärjestelmien käytön mahdollisuuksien nimissä niiden käyttöönottoa ja sovellutuksia halutaan lisätä.<sup>182</sup> Toisaalta politiikkadokumenttien ja sittemmin säädöksessä omaksutun lähestymistavan mukaan luotettavuus taataan ihmisvalvonnalla, jonka taustalla on ihmistoimijuuden säilyttäminen ja järjestelmän toimijuuden palauttaminen ihmistoimijaan.<sup>183</sup> Vaihtoehtoisia keinoja ihmisvalvonnalle ei ole esitetty ja tahto säilyttää ihmisvalvonta on omaksuttu jo politiikkadokumenteissa.

Kuten HCI-tutkimusta käsittelevässä jaksossa todettiin, ihmisvalvonnan heikkoudet on tunnistettu traditiossa jo 1980-luvulla.<sup>184</sup> Kritiikki ihmisvalvojan käyttöä ja ylipäätään hierarkiaa kohtaan on vain voimistunut. Parhaiden järjestelmien on todettu toimivan ilman kummankaan valta-asemaa. Valvonnan säilyttäminen liitettiin alkujaan osaksi funktioalokaation oletuksia, joiden mukaan ihminen on parempi toimimaan muuttuvassa ja monimutkaisessa ympäristössä ja järjestelmät puolestaan ennalta määrättyssä.<sup>185</sup> Tekoälysäädöksessä järjestelmiltä edellytetään toimintavarmuutta monimutkaisessakin ympäristössä.<sup>186</sup> Tästäkin syystä on kyseenalaista, missä määrin valvonnan vaatimusta voidaan teknisestä näkökulmasta perustella reaktiivisten tekoälyjärjestelmien kohdalla. Yksi HCI-tutkimuksen kritiikistä kohdistui automaattisten järjestelmien osalta kykyyn suorittaa tehtäviä huomattavalla nopeudella ihmiseen verrattuna, jolloin ihmisvalvoja ei kykene reaaliaikaiseen valvontaan. Nykyisen tekoälyteknologian kohdalla valvojalle haasteeksi muodostuu järjestelmän laskentatehon lisäksi päätöksenteon seuraamisen vaikeus läpinäkyvyyden ja adaptiivisuuden vuoksi. Tähän koitetaan toki puuttua säädöksen läpinäkyvyysvaatimuksilla.<sup>187</sup>

---

<sup>181</sup> Esim. Koulu 2020 s. 19-20.

<sup>182</sup> Tekoälysäädös johdanto-osan perustelukappaleet 3-5.

<sup>183</sup> Euroopan parlamentti 2019 kohta 62.

<sup>184</sup> Ks. Bainbridge 1983; Sheridan & Ferrell 1975.

<sup>185</sup> Laundry 2013 s. 778.

<sup>186</sup> Esim. johdanto-osan perustelukappale 75.

<sup>187</sup> Tekoälysäädöksen johdanto-osan perustelukappale 27 & artikla 13.

Muitakin HCI-tutkimuksen tunnistamia ihmisvalvonnan puutteita on koitettu paikata tekoälysäädöksessä. Esimerkiksi HCI-tutkimuksessa tunnistettua valvojan pätevydestä johtuvaa heikkoutta pyritään paikkamaan riittävällä koulutuksella ja tiedonsaannilla.<sup>188</sup> Lisäksi tekoälyjärjestelmiin pyritään sisään rakentamaan ihmisen valvonnan mahdollistamia toimintoja.<sup>189</sup> Järjestelmän tulee ohjeistaa ja tukea ihmistä hänen arvioidessa, milloin ihmisen suorittama interventio on tarpeellinen riskien tai negatiivisten seurausten välttämiseksi.<sup>190</sup> Huomion arvoista on, että HCI-tutkimus on lähestynyt samaa ongelmaa vastaavilla keinoilla ja siitä huolimatta ihmisvalvonnan tehokkuus on asetettu kyseenalaiseksi.<sup>191</sup> HCI-tutkimuksen muuttunut paradigma, jossa korostettavaa matalaa hierarkiaa ei näy nyt oikeudellistuvassa vuorovaikutuskäsityksessä. Oikeudellinen vuorovaikutuskäsitys on tältäkin osin dikotominen.

### 3.2.3 Järjestelmien toimijuus uhkana ihmisen toimijuudelle

Ihmistoimijuuden suojelun ja tukemisen tavoite yhdessä ihmisvalvonnan vaatimuksen kanssa on tehty säädöksessä eksplisiittisesti selväksi ja lähestulkoon kaikki säädöksen kohdat ovat palautettavissa ihmistoimijuuden suojelun tavoitteeseen. Jo käsitellyllä tavalla tämä ilmenee tekoälyjärjestelmien roolista ihmisen työvälineenä ja ihmisen päätöksenteon tukijana sekä ihmisen kontrollin viimesijaisuudesta. Ennen kuin analysoin, miksi tekoälysäädöksen toimijuuskäsitykset heijastavat nähdäkseni dikotomiaa osoitan, että HCI-tutkimuksen valossa on perusteltua puhua järjestelmien toimijuudesta.

Tekoälyn määritelmää koskevassa kappaleessa tekoälyn leimalliseksi ominaisuudeksi verrattuna sääntöpohjaisiin järjestelmiin (eng. “rule based systems”) liitetään sen kyky tehdä päätelmiä.<sup>192</sup> Tällä viitataan sen kykyyn tuottaa ennusteita, sisältöä,

---

188 Säädöksessä AI-lukutaito on käsitteellistys tekoälyjärjestelmien ymmärrykselle. Lukutaidon vaatimusta ilmentäviä kohtia on valtava määrä. Esimerkiksi johdanto-osan perustelukappaleen 72 mukaan korkean riskin järjestelmät suunniteltava siten, että käyttöönottajat voivat ymmärtää järjestelmän toimintaa, arvioida toiminnallisuutta ja ymmärtää vahvuudet sekä rajoitteet sen toiminnalle. Lisäksi korkean riskin järjestelmiin on laitettava käyttöohjeet (artikla (13(2)), joita käyttöönottajien on noudatettava (artikla (26(5))). Käyttöönottajien huolehdittava myös riittävästä teknisistä ja organisatorisista keinoista varmistaakseen, että järjestelmiä käytetään ohjeiden mukaan (26(1)) artikla.

189 Tekoälysäädös artikla 13(3(d)).

190 Tekoälysäädöksen johdanto-osan perustelukappale 73.

<sup>191</sup> Esim. Bainbridge 1983.

192 Tässä tehdään määritelmällisesti eroa autonomisten ja automaattisten järjestelmien välillä. HCI-tutkimuksen mukaan automaattisille järjestelmille ominaista on käskyjen noudattaminen

suosituksia tai päätöksiä, jotka vaikuttavat ympäristöön, jonka kanssa järjestelmä on vuorovaikutuksessa joko fyysisessä tai digitaalisesta ulottuvuudesta.

Tekoälyjärjestelmät toimivat vaihtelevilla autonomian asteilla. Osa järjestelmistä pystyy toimimaan jossain määrin itsenäisesti ilman ihmisen osallisuutta ja puuttumista.<sup>193</sup> Tekoälysäädöksen määritelmän mukaisesti järjestelmillä on kyky oppia itsenäisesti ja toimia adaptiivisesti. HCI-tutkimuksen mukaisesti autonomisuus on olennainen osa toimijuutta. Autonomisuus ei HCI-perinteen mukaan edellytä, että järjestelmällä on oma agenda vaan autonomiaa voi ilmentää myös pyrkimykset ennalta määrätyn tavoitteen saavuttamiseksi.<sup>194</sup> HCI-tutkimus puoltasi siten toimijuuden tunnustamista, vaikka säädöksessä ei nimenomaisesti puhuta autonomian ja ihmisestä riippumattomuuden yhteydessä järjestelmän toimijuudesta.

Etenkin tekoälysäädöksen riskien kuvailussa käytetty kieli synnyttää mielikuvia merkittävänkin autonomian olemassaolosta.<sup>195</sup> Lisäksi sekä tekoälyn määritelmä että osa tekoälyjärjestelmiä koskevista vaatimuksista tähdätään juuri niiden ominaisuuksien mitikointiin, joita HCI-tutkimuksessa tunnistettiin korkean autonomian järjestelmille. Esimerkiksi merkittävä läpinäkyvyyden painotus osoittaa, että järjestelmät käyttävät huomattavaa autonomiaa, sillä HCI-puolella läpinäkymättömyys eli päättelyketjun seuraamisen vaikeus liitetään korkeampaan autonomian asteeseen.<sup>196</sup> Nähdäkseni myös näiden syiden vuoksi HCI-tutkimuksen perusteella tekoälyjärjestelmillä on toimijuus. Säädöksessä oikeudellistuvaan vuorovaikutuskäsitykseen ei kuitenkaan sisälly ainakaan nimenomaisesti ilmaistua ajatusta järjestelmän toimijuudesta.

Näin ollen luonnollisesti mikään säädöksessä ei vaikuttanut viittaavan HCI-tutkimuksessa tunnistettuun me-toimijuuden mahdollisuuteen tai vuorovaikutuksen näkemiseen yhteisenä prosessina. Vaikutelman syntymiseksi järjestelmän implisiittisen toimijuuden lisäksi olisi olennaista pystyä tunnistamaan yhteisiä tavoitteita. Yhteisten tavoitteiden tunnistaminen on hankalaa tarkkarajaisten

---

deterministisesti ja kaavamaisesti sekä ja toteutus, joka koostuu suunnitellusta ja rajoitetusta joukosta ennalta määritellyjä tuloksia. Ks. Simmler & Frischknecht 2021.

193 Tekoälysäädöksen johdanto-osan perustelukappale 12 & artikla 3(1).

194 ks. Sterling, L. S., & Taveter, K. 2009.

195 Esimerkiksi tiettyjen yleiskäyttöisten tekoälymallien katsotaan muodostavan järjestelmätason riskin. Erityisesti tekoälysäädöksen johdanto-osan perustelukappaleet 110 & 111.

196 Toisin sanoen, mitä autonomisempi järjestelmä on, sen mahdottomampi ihmisen on seurata, miten ja miksi tietty syöte tuottaa tietyn tuloksen. Ks. Simmler & Frischknecht 2021.

tehtäväkuvausten vuoksi, joka johtuu järjestelmien roolista ihmisen apuvälineenä ja ihmistoimijan viimesijaisuudesta. Pikemminkin ihmisen ja järjestelmän toimijuus nähdään toisensa poissulkevinä, jolloin ihmistoimijuutta tulee suojella järjestelmän toimijuudelta. Ihmistoimijan tukemisen, säilyttämisen ja arvostamisen vaatimus on luotettavan tekoälyn edellytys ja suhde järjestelmien toimijuuteen onkin osin ristiriitainen. Tämä palautuu lopulta tutkielman alussa kuvattuun punnintaan järjestelmien käytön mahdollisuuksista ja riskeistä. Kaiken kaikkiaan käsitys vuorovaikutuksesta toimijuudenkin osalta on dikotominen.

### 3.2.4 Ihmisen asettamat erinäiset tavoitteet

Toimijuuden tavoin tavoitteiden osalta on ensin syytä tarkastella, voidaanko järjestelmällä käsittää olevan “omia” tavoitteita. Tekoälysäädöksessä on omaksuttu erottelu järjestelmien eksplisiittisten ja implisiittisten tavoitteiden välillä, joka käy ilmi tekoälyjärjestelmän määritelmästä.<sup>197</sup> Johdantokappaleen mukaan erottelulla on tarpeen korostaa, että tekoälyjärjestelmät voivat toimia nimenomaisesti määriteltyjen tavoitteiden tai implisiittisten tavoitteiden mukaan. Tekoälyjärjestelmän tavoitteet voivat tietyissä tilanteissa erota tekoälyjärjestelmän käyttötarkoituksesta.<sup>198</sup> Näin ollen selvää on, että järjestelmälle on asetettu jokin tavoite tai se toteuttaa jotain tavoitetta. Maininta siitä, että tavoite ei välttämättä ole sama kuin käyttötarkoitus merkinnee ainakin sitä, että tavoitetta ei aina määritellä järjestelmän käyttäjän kautta. Näin ollen järjestelmällä voi tekoälysäädöksen valossa katsoa olevan omia tavoitteita. Järjestelmille tekoälysäädöksessä asetettavat toiminnallisuusvaatimukset ovat mielestäni paras esimerkki järjestelmän tavoitteista. Järjestelmän tavoite on esimerkiksi toimia luotettavasti monimutkaisessakin ympäristössä.

Järjestelmälle osoitettava tavoite ei tarkoita samaa kuin järjestelmän määrittämä tavoite, vaikka välillä tekoälysäädöksen kielenkäyttö antaa ymmärtää, että järjestelmät itsessään pyrkisivät johonkin päämäärään.<sup>199</sup> Kuten HCI-tutkimuksessa tuotiin esiin autonomian yhteydessä, ei ole syytä edellyttää, että järjestelmä määrittäisi itse tavoitteet. Lähestymistapaa perustellaan HCI-tutkimuksessa

<sup>197</sup> Tekoälysäädös artikla 3(1).

<sup>198</sup> Tekoälysäädöksen johdanto-osan perustelukappale 12.

<sup>199</sup> Esimerkiksi kasvojen tunnistuksen yhteydessä erityinen huoli on kohdistettu tekoälyjärjestelmiin, jotka pyrkivät tunnistamaan tai päättelemään tunteita. Ks. tekoälysäädöksen johdanto-osan perustelukappale 44. Toisin kuin moni muu kohta tekoälysäädöksessä, tämä esittää pyrkimyksen olevan järjestelmällä itsellään. Kuitenkin myöhemmin samassa kappaleessa todetaan järjestelmien käytön (eng. “intended to be used to detect”) olevan kiellon kohteena.

tarkoituksenmukaisuudella. Olisi tuskin käytännöllistä, jos ihmisen luomien järjestelmien tulisi toimia oman agendan mukaan ja ihmisen tarkoituksen vastaisesti, jotta niitä voitaisiin pitää autonomisina toimijoina. Tavoitteet voivat olla siten ulkopuolelta asetettuja. Tässä mielessä järjestelmät ymmärretään HCI-tutkimuksen puolella edustajina.<sup>200</sup> Näin ollen järjestelmillä voi sekä tekoälysäädöksen että HCI-tutkimuksen käsityksen mukaan olla omia tavoitteita.

Yhteinen tavoite olisi puolestaan ymmärrettävä päämääräksi, jota kohden molemmat työskentelevät. Yhteisten tavoitteiden nimeäminen tekoälysäädöksessä on hankalaa johtuen tarkkarajaisista tehtävän kuvauksista ja järjestelmille varatuista avustavien osatehtävien suorittamisesta. Tämä ei ole yhteensopiva HCI-tutkimuksen tarkoittaman yhteistyömallin kanssa. Yhteisten tavoitteiden tulisi näkyä ajatuksena, jonka mukaan vuorovaikutus on enemmän kuin osiensa summa. Tämä ei toteudu, koska järjestelmät nähdään ennen kaikkea ihmisen työvälineenä, HCI-tutkimuksen valitseman päämieslähestymistavan sijasta.

Tekoälysäädöksessä ei ole havaittavissa kielenkäyttöä, jonka mukaan järjestelmän tehtävät olisi ymmärrettävä järjestelmän tavoitteiksi toiminnallisuuksien tapaan. Johtuen järjestelmien luonteesta työvälineenä oikeudellisessa vuorovaikutuskäsityksessä olennaista on enemmänkin tarkoitus, johon järjestelmää käytetään. Esimerkiksi tekoälyjärjestelmien valmistelevia tehtäviä korostavista kohdista voidaan todeta, että tiettyjen suurempien tavoitteiden tai tehtäväkokonaisuuksien hoitaminen nähdään tekoälyjärjestelmien käytön tavoitteeksi. Tehtäväkokonaisuuksien hoitamista ei hahmoteta ihminen-järjestelmä yhteenliittymälle asetetuksi vaan käyttäjälähtöisesti, joka täyttää oman tavoitteensa hyödyntämällä järjestelmää apuvälineenä. Esimerkiksi biometriseen tunnistukseen käytettäviä järjestelmiä hyödynnetään lainvalvonnalle asetettujen tavoitteiden täyttämiseksi.<sup>201</sup> Käyttäjän avulla määritettävät tehtävät ovat hankala mieltää yhteisiksi eli ihminen-järjestelmä vuorovaikutuksen tavoitteiksi muiden tunnistettujen vuorovaikutusmuuttujien vuoksi. Tekoälysäädöksessä omaksutun tarkkarajaisen roolituksen, joka perustuu eritellyille osatehtäville, lisäksi etenkin hierarkia puhuu yhteisiä tehtäväkokonaisuuksia vastaan.

---

<sup>200</sup> Sterling & Taveter 2009.

<sup>201</sup> Tekoälysäädöksen johdanto-osan perustelukappale 54.

Siten tekoälysäädöksessä ei ole havaittavissa kohtia, joidenka mukaan järjestelmän tehtävät olisi ymmärrettävä sen enempää järjestelmän kuin ihminen-järjestelmä yhteenliittymän tavoitteiksi. Yllä koitin kuitenkin kuvata, kuinka järjestelmältä vaadittavat ominaisuudet voidaan mieltää järjestelmän omiksi tavoitteiksi, vaikka ne olisivatkin ihmisen järjestelmille asettamia.<sup>202</sup> Yksinkertaistettuna on selvää, että oikeudellistuvassa vuorovaikutuskäsityksessä ei ole havaittavissa yhteisiä tavoitteita ihmiselle ja järjestelmälle. Näin ollen riippumatta siitä, hyväksytäänkö edes järjestelmän ominaisuudet sen omiksi tavoitteiksi, oikeudellistuvan vuorovaikutuskäsityksen dikotominen luonne ei muutu.

### 3.3 Symbioottisen vuorovaikutuskäsityksen häivähdykset

Kuvasin edellä järjestelmien toiminallisuuksien ja osatehtävien eron tarkkaan, koska haluan esittää mahdollisuuden mieltää järjestelmän toiminnallisuudet ja tekoälyjärjestelmien käyttäjälle asetetut vaatimukset yhteisenä päämääränä ja siten häivähdyksenä symbioosista.<sup>203</sup> Huolimatta erillisistä tavoitteista, toiminnallisuudet ja käyttöönottajalle asetetut vaatimukset voivat kohdistua samaan päämäärään. Esimerkiksi järjestelmälle asetettu vaatimus tulosten tulkittavuudesta ja vastaavasti ihmiselle asetettu vaatimus AI-lukutaidosta palvelevat molemmat valistunutta tekoälyjärjestelmän käyttöä. Samoin järjestelmältä edellytetään toiminnallisuutena käyttöliittymää, jonka tulisi mahdollistaa ihmisen suorittama valvonta. Myös muut järjestelmille asetetut vaatimukset yhdessä ihmisten velvollisuuksien kanssa kohdistuvat säädöksessä esiintuotuihin päämääriin. Päämäärät ovat toki ihmisen asettamat sekä määrittelevät. Tämä ei kuitenkaan estä ajatusta symbioosista. Lickliderinkin symbioosimallissa lähtökohtana oli se, että ihminen asettaa päämäärät ja järjestelmälle annetaan taasen mahdollisuus toteuttaa näitä (eng. “do the scientific thinking”).<sup>204</sup>

---

<sup>202</sup> Nähdäkseni parhaiten tätä kuvaa vaatimus, jonka mukaan järjestelmän edellytetään toimivan huolimatta ihmisen mahdollisesta käyttötarkoituksen vastaisesta käytöstä. Esim. tekoälysäädöksen artikla 15(3). Toisin sanoen, vaikka ihminen toimisi järjestelmän tavoitetta vastaan. Tämä on mielestäni osoitus siitä, että toiminnallisuudet muodostavat järjestelmien omat tavoitteet, koska ihminen voi toimia niiden vastaisesti. Ajatus vuorovaikutuksen vahingollisuudesta saa aikaan käsityksen erillisistä tavoitteista, sillä harvoin toimijoiden oletetaan toimivan omia tavoitteitaan vastaan. On tuskin tarkoituksenmukaista olettaa, että ihminen toimisi vastoin omia tavoitteita, ja siksi miellän nämä järjestelmälle omaksi.

<sup>203</sup> Vaatimuksia ovat esimerkiksi läpinäkyvyys ja toimintavarmuus. Tekoälysäädöksen johdanto-osan perustelukappale 27.

<sup>204</sup> Licklider 1960 s. 4-11.

Ongelmaksi symbioottisen vuorovaikutuskäsityksen mahdollisuudelle tältäkin osin muodostuu kuitenkin tekoälysäädöksessä omaksuttu hierarkia. Ihmiskeskeisen tekoälyn tulisi toimia työvälineenä ihmiselle ja sen tekoälyn perimmäisenä tarkoituksena tulee olla ihmisen hyvinvoinnin lisääminen.<sup>205</sup> Nähdäkseni tekoälyjärjestelmät tulisi HCI-tutkimuksen mukaisesti ensin mieltää edustajiksi eikä työvälineiksi, jotta yhteistyösuhteesta olisi mielekästä puhua. Koska järjestelmät ovat vain työvälineitä, ihmisen ja järjestelmän vuorovaikutuksen aste oikeudellisesti HCI-käsitteitä käyttäen on yhteiselon (eng. “co-existence”) tasolla eli ne jakavat saman työtilan.<sup>206</sup> Oikeudellinen vuorovaikutuskäsitys on näin ollen jäljessä HCI-tutkimuksen vuorovaikutuskäsityksen asteesta, jossa korostetaan yhteistyön merkitystä.

Tiivistetysti voin todeta, että ne kohdat, jotka voisivat muutoin ilmentää vastavuoroisuutta, joustavuutta ja matalan hierarkian yhteistyötä muuttuvat arvottaviksi ja valta-asetelman paljastaviksi semantiikan tai riskiluokitusystematiikan vuoksi. Kohdittain lineaarisesta työnjakamisesta ja arvottamisesta etäännyttään hieman ja vuorovaikutus saa sirkulaarisia piirteitä. Ihmisen ja järjestelmän vuorovaikutuksessa järjestelmä voi esimerkiksi joissain tilanteissa suorittaa ihmistyön päälle tehtäviään. Ihminen vastaavasti lisää tähän vielä oman kerroksen ja näin syntyy kehämäistä vuorovaikutusta.<sup>207</sup> Ihminen säilyy kuitenkin poikkeuksetta kontrollissa ja vallan käyttäjänä, jolloin dikotominen vuorovaikutusmalli pysyy vallitsevana.

Symbioottisen suhteen mahdollisuuden puolustamiseksi on myös syytä muistuttaa tekoälysäädöksen riskiperusteisesta lähestymistavasta.<sup>208</sup> Dikotomista vuorovaikutusta eniten alleviivaavat vaatimukset ovat asetettu korkean riskin tekoälyjärjestelmille, ja siten sääntely mahdollistaa symbioottiselle suhteelle perustuvan järjestelmäkehityksen matalan riskin tekoälyjärjestelmissä. Järjestelmät,

---

205 Tekoälysäädöksen johdanto-osan kappale 6. Tavoite on muutoinkin kirjoitettu mielenkiintoisesti, sillä siinä edellytetään tekoälyjärjestelmän lisäävän ihmisen hyvinvointia, vaikka järjestelmä mielletään vain ihmisen apuvälineeksi. Kohta ei huomioi vuorovaikuttavan ihmisen osuutta tavoitteen toteuttamisessa, vaikka apuväline sana viittaa siihen, että ihmisellä on oltava aktiivinen rooli tavoitteen toteuttamisessa. Joko tekoälysäädöksessä ei ole huomioitu tätä sanavalintojen epäjohtonmukaisuutta tai siinä implisiittisesti asetetaan ihmistoimijan automaattiseksi tavoitteeksi hyvinvoinnin lisääminen, jolloin siitä voisi muodostua yhteinen tavoite ihmisen ja järjestelmän vuorovaikutukselle.

206 Schmidtler et al. 2015 s. 83-95.

207 Tekoälysäädöksen johdanto-osan perustelukappale 53.

208 Euroopan komissio COM(2020) 65 final s. 19.



jotka jäävät tekoälysäädöksen seitsemän vaatimuksen soveltamisen ulkopuolelle, ovat ensisijaisesti muun alussa tunnistetun sääntelyn viitekehyksessä ja siten tämän tutkielman ulkopuolella, jossa keskiössä on nyt oikeudellistuva eli tekoälysäädöksen ilmentämä vuorovaikutuskäsitys.

### 3.4 Lopuksi

Kolme neljästä vuorovaikutusmuuttujasta osoittaa oikeudellistuvan vuorovaikutuskäsityksen olevan selkeästi dikotominen. Vain tavoitteiden osalta voidaan eritellä muita kuin selkeän dikotomisista esimerkkejä. Ottaen huomioon tekoälysäädöksessä omaksutut lähtökohdat luotettavan tekoälyn rakentamiselle ja muiden vuorovaikutusmuuttujien vaikutus, ei tavoitteitakaan voida nähdä puhtaasti symbioottisen käsityksen ilmentäjänä. Oikeudellistuva vuorovaikutuskäsitys on siten ilmeisen dikotominen.

Kuten HCI-tutkimusjaksossa todettiin, kerran juurtuneella teoriolla voi olla kauaskantoisia vaikutuksia. Dikotomisella vuorovaikutuskäsityksellä on ollut juuri tällainen vaikutus. Huomaamme käsityksen nyt myös oikeudellistuvan tekoälysäädöksen myötä, vaikka HCI-traditio on todennut dikotomisen lähestymistavan olevan etenkin autonomisten järjestelmien kanssa vuorovaikuttamisen mallintamiseen riittämätön. Oikeudellistuva vuorovaikutuskäsitys ei siten sovi yhteen teknologian kehityksen asteen kanssa.

HCI-tutkimus totesi ihmisen osuuden vuorovaikutusta ohjaavana tekijänä vähenevän kokonaan tai merkittävästi autonomisten järjestelmien kanssa vuorovaikutuksessa. Nyt oikeudellistuva vuorovaikutuskäsitys perustuu olennaisilta osin ihmisen rooliin vuorovaikutusta ohjaavana toimijana. Käsitys oli HCI-tutkimuksessa vallalla 60–80-luvuilla ja liittyi automaattisten järjestelmien kanssa vuorovaikuttamiseen. Sen haasteellisuus huomattiin jo näinä suosion päivinä. HCI-tutkimuksen kritiikki tiedostetaan osittain valvonnan haasteiden osalta, sillä tekoälysäädös pyrkii sääntelyllään valvonnan mahdollistamaan suuntaan esimerkiksi läpinäkyvyyden vaatimusten avulla. Säädöksessä kuitenkin vaietaan muilta osin ihmisen puutteellisista kyvyistä valvontaan. Valittaessa lähestymistapaa oikeudellistuvaan vuorovaikutuskäsitykseen, tämä on joko tietoisesti tai tiedostamatta jätetty noteeraamatta.

Samaan tapaan kuin tekoälysäädöksessä HCI-tutkimus esittää ihmisten ja autonomisten järjestelmien vuorovaikutuksen haasteeksi luottamuksen kokemisen. Myös tunnistetut haasteet luottamuksen rakentumiselle ovat yhtäläiset. Niitä ovat autonomisten järjestelmien yllätyksellisyys, arvoituksellisuus ja dynaamisuus. Luottamuksen rakentamisen keinot ovat kuitenkin joiltain osin toisistaan poikkeavat. HCI-tutkimuksessa esitetään, että luottamuksen rakentamiseksi suunnittelun tulisi jäljitellä ihminen-ihminen vuorovaikutuksessa havaittuja parhaita käytäntöjä. Oikeudellistuvassa vuorovaikutuskäsityksessä luottamuksen rakentamisen edellytys on ihmisen suorittama valvonta ja hierarkian säilyttäminen.

Mielenkiintoisesti sekä HCI-tutkimuksessa että oikeuden kontekstissa pyritään kuitenkin samaan ihanteeseen, joka on arvojen ja normien sisään rakentaminen järjestelmiin luottamuksen varmistamiseksi. Oikeudellinen ihmiskeskeisyys tarkoittaa järjestelmien mieltämistä työvälaineiksi, joka palauttaa automaattisesti toimijuuden ihmiseen varjellen ihmisen päätäntävaltaa ja kontrollia suhteessa järjestelmiin. Tällaiselle ihmiskeskeisyydelle rakentuu oikeudellinen vuorovaikutuskäsitys, joka sisään rakennetaan esimerkiksi mahdollistamalla ihmisen kontrollikeinot.<sup>209</sup> Ihmiskeskeisyyden lähestyminen suunnittelun näkökulmasta maalailee hyvin erilaisen kuvan ihmiskeskeisyyden merkityksestä. Keskiössä on autonomisten järjestelmien rakentaminen osaksi yhteisöjä ja ympäristöjä, jossa ne toimivat. Ihmisen toimijuus ei korostu suhteessa järjestelmiin, vaan keskiössä on vuorovaikutuksen onnistumisen arviointi. Arvojen implisiittisen sisään rakentamisen onnistumista arvioidaan järjestelmän suoriutumisen kautta, jos järjestelmä ei toimi osana ympäristöä, normien vastaavuutta ei ole.<sup>210</sup>

Luottamuksen rakentamisen keinot vuorovaikutuksessa vaikuttavat siten olevan perustavanlaatuisella tasolla erilaiset, vaikka ihanne ja pyrkimys ovat samat. Nyt kun tutkimuskysymyksiin yksi ja kaksi on vastattu eli oikeudellistuva vuorovaikutuskäsitys ja sen suhde HCI-tutkimuksen vuorovaikutuskäsitykseen on kuvattu, siirryn käsittelemään kolmatta tutkimuskysymystä eli keskustelemaan oikeusjärjestelmän rajoitteista erojen ilmentymisen selittäjänä ja toisaalta tuomalla esiin, mitä teknologiasäätelylle ominaisia haasteita erot ilmentävät.

---

209 Vrt. tekoälysäädöksen artikla 14.

210 Vrt. IEEE 2018 s. 33-34.

## 4 Oikeuden digitalisaatio ja teknologiasääntelyn haasteet

Seuraavaksi tarkastelen aineiston analyysin tuloksia oikeusjärjestelmän näkökulmasta ja liitän tulokset oikeuden digitalisaatiosta käytävään keskusteluun. Tarkoituksena on selittää tuloksia arvioimalla, miltä osin oikeusjärjestelmämme rakenteet pakottavat dikotomista vuorovaikutusmallia. Jaksossa viisi käsittelen niitä oikeuden lähtökohtia, jotka selittävät valittuja lähestymistapoja - eli ihmisvalvontaa, ihmistoimijuutta ja ihmisen suorittamaa päätöksentekoa - luottamuksen rakentamiselle sääntelyssä, ja estävät HCI-tutkimuksessa paremmaksi todetun symbioottisen vuorovaikutussuhteen luomisen.

Aloitin esittelemällä tekoälysäädöksestä havaittuja teknologiasääntelyn haasteiden ilmentymiä, jotka palautuvat oikeuden digitalisaation ilmiöön. Oikeuden, sääntelyn ja teknologian viitekehystä edustavassa tutkimuskirjallisuudessa on tunnistettu eri tasoja, joilla teknologiat ja niiden sääntely asettavat haasteita oikeudelle ja yhteiskunnalle.<sup>211</sup> Näitä tasoja on kolme ja ne voidaan jaotella sen mukaan, kuinka perustavanlaatuisen arvoon, instituutioon tai käytäntöön vaikutus kohdistuu. Ensimmäisenä voidaan mainita teknologioiden haasteet normatiivisille arvoille ja ideaaleille, jotka muodostavat oikeusjärjestelmämme perustan. Toiseksi voimme mainita oikeusjärjestelmämme rakenteille, käsitteille ja instituutioille aiheutuvat haasteet. Viimeisellä tasolla on haasteet tarkoituksenmukaisen sääntelykehyksen luomisessa, joka kestäisi aikaa.<sup>212</sup>

Tiedostan, että päätökseni olla rajaamatta analyysin tulosten oikeudellista käsittelyä yksinomaan ihmisvalvontaan, ihmisen toimijuuteen (järjestelmä työvälineenä) tai ihmisen suorittamaan päätöksentekoon aiheuttaa sen, ettei jaksossa voida yksittäisen vaatimuksen osalta käsitellä aihetta hirveän syvällisesti. Koska oikeudellinen vuorovaikutuskäsitys on kokonaisuus kaikkea näitä kolmea toisiinsa tiiviisti linkittyvää vaatimusta, katson kuitenkin tärkeämmäksi yleisen tason kuvailun. Lisäksi tarkoitukseni ei ole esittää ratkaisuja oikeuden digitalisaation tai

---

<sup>211</sup> Kansainvälisessä kirjallisuudessa esitetty jaottelu haasteiden paikantamisesta muistuttaa kotimaisessa oikeuskirjallisuudessa paljon hyödynnettyä oikeuden taso ajattelua. Oikeuden pintataso on ajassa ja paikassa nopeasti muuttuva, mutta oikeuden syvärakenteet ovat eri oikeuden alat läpäiseviä ja juurtuneita sekä aikaa kestäviä. Ks. esim. Tuori 2013 s. 31-42.

<sup>212</sup> Brownsword et al. 2017 s. 3-38.

teknologiasääntelyn haasteisiin vaan kuvata, mitä haastetta vuorovaikutuskäsityksen yksittäiset tekijät ilmentävät ja miksi sääntely on haastavaa.

## 4.1 Oikeuden antroposentrismi ja oikeuden digitalisaatio

### 4.1.1 Ihmiskeskeinen oikeus

Koko oikeusjärjestelmämme rakentuu ajatukselle ihmistoimijasta. Oikeuden voidaankin sanoa olevan korostuneen antroposentristä.<sup>213</sup> Lainsäädäntömme kohteena on nimenomaan ihminen ja lait muodostavat kehikon ihmisten väliselle vuorovaikutukselle.<sup>214</sup> Sääntelyn substanssin lisäksi oikeudelliset termit ja käsitteet kuvaavat ominaisuuksia, joita vuorovaikutajiin liitetään. Valittujen termien avulla on mahdollista luoda tietynlaista kuvaa vuorovaikuttajista tai tuoda ilmi suoranaisia valtasuhteita.<sup>215</sup> Termien ja käsitteiden tiedostamaton tai tiedostettu käyttäminen muovaa siten käsitystämme siitä, millaisena vuorovaikutus hahmotetaan.<sup>216</sup>

Teknologian kehitys on tuonut muutoksen vuorovaikutussuhteisiin. Ihminen ei ole ainoa vuorovaikutussuhteen mahdollinen osapuoli. Ihmistoimijuuden lisäksi voimme puhua esimerkiksi algoritmisesta toimijuudesta ja päätöksenteon osalta ihmisen ja järjestelmän suorittamasta hybridipäätöksenteosta.<sup>217</sup> Haaste, jonka oikeus kohtaa, on päätöksenteon ja toimijuuden etääntyminen ihmisestä. Ihmisen ja toiminnan välille kerrostuu algoritmista toimijuutta ja näin ollen etäisyys ihmisen ja toiminnan seurauksen välillä kasvaa.<sup>218</sup>

Oikeudellisella sääntelyllä on kuitenkin totuttu vaikuttamaan nimenomaan ihmisen käyttäytymiseen. Ihminen toimii välittäjänä sääntelyn kohdistuessa esimerkiksi luontoon tai teknologiaan. Ihminen on ollut luonnollinen valinta sääntelylle, koska vain ihmisellä on tarvittava kyky vaikuttaa oikeudellisista interventioista.<sup>219</sup>

Tekoälysäädös ei tuo muutosta oikeuden antroposentrismiin. Kuten jo politiikkadokumenteissa todettiin, tekoälysäädöksellä luodaan ihmisen

---

<sup>213</sup> Esim. Koulu 2018 s. 16; Viljanen 2020 s. 1071.

<sup>214</sup> Viljanen 2020 s. 1071-1072; Wein 1992 s. 103-105.

<sup>215</sup> Ks. esim. Koulu et al. 2022 s. 14-15. Artikkelissa nostetaan esiin oikeudellisten käsitteiden muutos digitalisaatiokontekstissa. Käsitteillä on merkittävä ero arvioitaessa toimijaan liitettävää aktiivisuuden ja passiivisuuden tasoa.

<sup>216</sup> Ks. Koulu et al. 2024.

<sup>217</sup> Ks. esim. Koulu 2020 s. 38.

<sup>218</sup> Viljanen 2020 s. 1071.

<sup>219</sup> Viljanen 2020 s. 1071. Viljanen puhuu ihmisen kyvystä vaikuttaa oikeudellisista interventioista. Kyky vaikuttaa interventiosta johtaa siihen, että sääntelyllä voidaan muuttaa maailmaa siitä, mikä se olisi ollut ilman interventiota.

käyttäytymistä säänteleviä sosiaalisen valvonnan mekanismeja.<sup>220</sup> Ihminen on siten yhä sääntelyn kohteena. Näin ollen voimme arvella tekoälyjärjestelmien toimijuuden ja tekoälysäädöksen ihmisiin kohdistuvien vaatimusten aiheuttavan haasteita sääntelyn tarkoituksenmukaisuudelle. Kuten aineiston analyysin yhteydessä esitettiin, oikeudellinen ihmiskeskeisyys tarkoittaa päätöksenteon ja toimijuuden palauttamista ihmiseen.

#### 4.1.2 Haasteiden käsitteellistäminen: Oikeuden digitalisaatio

Muuttuviin vuorovaikutussuhteisiin on kiinnitetty huomiota oikeustieteessä jo pitkään. Yhdysvaltalainen oikeustieteilijä Wein on todennut jo 1992, että 2000-luvulla “digitaalinen aikakausi kohtaa lain todellisuudessa, jossa ihmisen ja tietokoneen välinen vuorovaikutus on pikemminkin sääntö kuin poikkeus”.<sup>221</sup> Hän on todennut myös, ettei pidä olla ihmeissään, jos “-- tuomarit insinöörien sijaan määrittävät suunnitteluprosessia enemmän kuin insinöörit uusien teknologioiden osalta”. Weinin mukaan oikeusjärjestelmän rooli vuorovaikutuksessa tulee oleman välimiesmäinen ja lakien avulla pidetään teknologia- “monsteri” aisoissa.<sup>222</sup> Kommenteista on luettavissa vaatimus reagoida vuorovaikutussuhteissa tapahtuvaan muutokseen oikeudellisesti sekä vahva usko lainsäädännön rooliin muutoksen suitsijana ja hillitsijänä. Kolmekymmentä vuotta myöhemmin voimme todeta Wienin olleen oikeassa ainakin muutoksen laajuudesta ja sääntelytarpeesta. Myös ajatus sääntelyn kohdistamisesta suunnitteluprosessiin on ansiokas, joskaan ei ole toteutunut yhtä ilmeisenä kuin Wien on esittänyt, vaikka tekoälysäädös onkin suunnattu pitkälti järjestelmien kehittäjille.<sup>223</sup>

Digitalisaatio käsitteenä viittaa yhteiskunnalliseen murrokseen, joka on syntynyt tieto- ja viestintäteknologioiden käytön yleistymisestä ja lisääntyvästä hyödyntämisestä.<sup>224</sup> Digitalisaation käsitteen sijaan voidaan alleviivata teknologian kehityksen astetta ja eritellä ilmiötä tarkemmin teknologian ominaisuuksia kuvaavien termein. Voidaan puhua esimerkiksi laskennallistumisesta, tietokoneistumisesta ja

---

220 Euroopan komissio COM(2020) 65 final s. 13.

221 Wein 1992 s. 105.

222 Ibid. s. 104.

223 Itseasiassa roolien kääntymisestä on esitetty päinvastaista “by design”-sääntelyn osalta. Järjestelmäsuunnittelijoiden tulisi kyetä sisään rakentamaan oikeudellisia tavoitteita. Samalla jälkikäteisestä oikeudesta siirrytään ennakkolliseen. Ks. esim. Urquhart & Rodden 2016.

224 Koulu 2018 s. 843-844.

tekoälyllistymisestä.<sup>225</sup> Digitalisaatio ei tuo käyttöömme ainoastaan uutta teknologiaa vaan se muovaa lisäksi sosiaalisia käytäntöjä. Digitalisaation tuoma muutos on siten myös laadullinen.<sup>226</sup>

Oikeuden digitalisaatioon on havahduttu yhteiskunnan digitalisaation kautta. Oikeustieteellinen keskustelu digitalisaation vaikutuksista ja mahdollisuuksista ei ole vasta taannoin syntynyttä. Oikeuden ja teknologian suhteen pioneerina toiminut Loevinger on jo 1949 oikeusinformatiikaksi (eng. “jurimetrics”) nimetyllä oikeudenalalla hahmotellut, kuinka laskennallisia tekniikoita voisi käyttää oikeudellisessa päättelyssä.<sup>227</sup> Loevinger kohdisti kritiikkiä oikeuden näennäisenä näkemäänsä monimutkaisuuteen ja oikeudellisen päätöksenteon epäloogisuuteen.<sup>228</sup> Muita tunnettuja digitalisaatiota tutkineita oikeustieteilijöitä on esimerkiksi Lawrence Lessig, joka teki kuuluisan vertauksen koodin ja oikeuden normatiivisuuden välillä. Koodin normatiivisuudella hän tarkoitti sääntelyn muotoa, jossa ohjelmistosuunnittelijat määrittävät arvojensa mukaiset vuorovaikutuksen säännöt, jotka “kyberavaruudessa” tulevat laajemmin noudatetuksi.<sup>229</sup> Oikeuden ja teknologian suhteen teorioinneissa on siirrytty korostamaan kummankin normatiivisen luonteen sijaan niiden keskinäistä dynaamista vuorovaikutussuhdetta.<sup>230</sup>

Oikeuden digitalisaatio tarkoittaa oikeuden sisäisestä näkökulmasta niitä moninaisia tapoja, joilla teknologiat muovaavat lainkäyttöä, lainsäädäntöä ja oikeustiedettä eli oikeudellisia käytäntöjä.<sup>231</sup> Oikeuden digitalisaatio tarkoittaa siten paljon muutakin kuin uusia aineellisoikeudellisia kysymyksenasetteluita. Vaikutukset ovat horisontaalisia, oikeudenalarajat ylittäviä ja pakottavat kyseenalaistamaan vallitsevassa oikeusjärjestelmässämme perusteina olevia käsitteitä ja yleisiä oppeja.<sup>232</sup> Esimerkkejä oikeudellisten käytäntöjen digitalisaatioista voidaan hakea suoraan tekoälysäädöksestä. Tekoälysäädöksessä määritellään esimerkiksi, missä määrin tekoälyjärjestelmiä voidaan hyödyntää lainkäytössä oikeudellisen

---

225 Sankari et al. 2022 s. 185.

226 Koulu 2018 s. 843-844.

227 Ks. Loevinger 1949 s. 455-493.

228 Ibid.

229 Lessig 1999 s. 4-8.

230 Ks. esim. Hildebrandt 2019 & Cohen 2019.

231 Esim. Koulu 2018 s. 841-843. Ks. lisää oikeudellisista käytännöistä Tuori 2003.

232 Ks. Bennet Moses 2013 s. 7.

päätöksenteon apuna. Oikeudellisen päätöksenteon kontekstissa järjestelmiä voidaan käyttää hallinnollisten tehtävien suorittamiseen, kun järjestelmät eivät vaikuta itse asiassältöön.<sup>233</sup> Muutokset lainsäädäntöön ilmenevät esimerkiksi aineellisoikeudellisella tasolla ja tekoälysäädös on itsessään osoitus uudesta aineellisesta lainsäädännöstä. Tekoälyteknologialla nähtiin olevan siinä määrin erityispiirteitä, että oli tarpeen luoda tekoälyteknologiaspesifiä sääntelyä.<sup>234</sup> Tekoälysäädökseen omaksuttu tarve suojella ihmistoimijaa puolestaan osoittaa, että tekoälyjärjestelmän toimijuuden katsotaan haastavan ihmistoimijan ja sen myötä vaikuttavan keskeisiin oikeudellisiin käsitteisiinmekin.<sup>235</sup> Muutos oikeustieteen tutkimukselle voi näkyä esimerkiksi empiirisen tutkimuksen lisääntymisenä tai teknologian hyödyntämisenä tutkimuksessa. Toisaalta oikeustieteilijöiden keskuudessa korostettu tarve monimetodisen oikeudellisen tutkimuksen tekemiselle osoittaa teknologian muovaavan myös tutkimuksen painotuksia.<sup>236</sup>

## 4.2 Teknologiasääntelyn haasteet

### 4.2.1 Teknologiasääntelyn tutkimuksen suhde sääntelyyn

Tutkimuskirjallisuudessa on huomautettu, kuinka teknologiasääntelyn tutkimus (eng. “technology regulation”) on suuri ja epämääräinen viitekehys, joka pitää sisällään ison määrän erilaista tutkimusta. Teknologiasääntelyn tutkimuksen viitekehyksessä kysymyksenasettelut ovat usein tarkkarajaisia. Näin ollen tutkimus siiloistuu ja kokonaiskuvan muodostuminen vaikeutuu.<sup>237</sup> Onkin hyvä huomata, että uuden teknologian tutkimus ei välttämättä ole teknologiasääntelyn tutkimusta vaan sääntelyn tutkimusta uuden teknologian valossa.<sup>238</sup> Kuten oikeudellisten käytäntöjen digitalisaatiosta esitettyjen esimerkkien avulla havainnollistin, ilmiö on moninainen ja siksi vaatisi tarkkarajaisen kysymysten lisäksi holistista lähestymistapaa.<sup>239</sup>

---

<sup>233</sup> Tekoälysäädöksen johdanto-osan perustelukappale 53 ja artikla 6(3(d)).

<sup>234</sup> Euroopan komissio COM(2020) 65 final s. 11.

<sup>235</sup> Morse 2017 s.153 “ -- are likely to transform the law’s traditional concepts of the person, agency, and responsibility.”

<sup>236</sup> Esim. Bennet Moses 2013; Brownsword 2016. Oikeuden ulkoisesta näkökulmasta ilmiötä voidaan tarkastella laajemmin ja pohtia oikeuden digitalisaation rajoja. Tällainen ulkopuolinen näkökulma voisi esimerkiksi pyrkiä selvittämään, miten uudet teknologiaa hyödyntävät oikeudelliset käytännöt vaikuttavat yhteiskunnassa. Sankari et al. 2022 s. 187-188.

<sup>237</sup> Bennet Moses 2013 s. 1-4.

<sup>238</sup> Ibid. s. 3.

<sup>239</sup> Esim. Sankari et al. 2022 & Bennet Moses 2013.

Teknologiasääntelyn tutkimuksella on luonnollisesti vaikutusta siihen, millaisia lähestymistapoja teknologioiden sääntelyyn valitaan. Teknologiasääntelyn tutkimuksen haasteet näyttäytyvät siten osin samaistettavina teknologiasääntelyn haasteina. Esimerkiksi yksi teknologiasääntelyn haasteista on sektorikohtainen lainsäädäntö. Sääntelyn pohjaksi olisi hyvä olla laadukasta, oikeudenala- ja tieteenalarajat ylittävää tutkimustietoa erityisesti tehokkuuden ja legitiimiyden varmistamiseksi.<sup>240</sup> Huolimatta sääntelyn haasteellisuudesta, se on ollut pitkään tapa vaikuttaa teknologioihin ja toisaalta ohjata oikeuden digitalisaatiota. Erilaisten teknologioiden sääntelyä on ollut kautta historian, eikä viimeaikaiset datavetoiset teknologiat ole näin ollen totaalinen poikkeus.<sup>241</sup> Tekoälysäädöksen kunnianhimoisista tavoitteista huolimatta on kuitenkin hyvä muistaa, että säädös ei ole teknologianeutraali.<sup>242</sup> Se soveltuu vain säädöksessä määriteltyihin tekoälyteknologioihin ja jättää näin ollen esimerkiksi sääntöpohjaisen automaation soveltamisalan ulkopuolelle.<sup>243</sup> Siten tekoälysäädös toistaa haasteeksi tunnistettua sektorikohtaisuutta. Sen sijaan, että sääntelyn kohteeksi valittaisiin tietty teknologia, sääntelyllä voitaisiin pyrkiä vaikuttamaan siihen sosiotekniseen käytäntöön, joka teknologian käytöstä aiheutuu ja jonka tunnistetaan muodostavan riskin oikeuksien toteutumiselle.<sup>244</sup> Siten sääntely voisi kohdistua esimerkiksi muun toimijan kuin ihmisen suorittamaan päätöksentekoon yleisesti.

#### 4.2.2 Teknologian käsitteestä

Tutkimuskirjallisuudessa on koitettu selventää sekä sääntelyn että teknologian käsitteitä. Näistä sääntelyn käsitteeseen kohdistetaan merkittävästi enemmän huomiota.<sup>245</sup> Selkeästi harvemmin perustellaan niin sanotun teknologiasääntelyn kohteena olevan teknologian tutkimista tai määritellään teknologia nimenomaisesti.

---

<sup>240</sup> Sankari et al. 2022 s. 189.

<sup>241</sup> Ibid.

<sup>242</sup> Grady 2023.

<sup>243</sup> Tekoälysäädös artikla 3(1). Nähdäkseni tekoälyteknologianeutraalius voidaan myös kyseenalaistaa, sillä trilogineuvottelujen lopputuloksena hyväksytyyn versioon sisällytettiin yleismalleja koskevat erityiset horisontaaliset säännöt. Ks. Euroopan komissio (2024) Interinstitutional File: 2021/0106(COD), Main elements of the compromise kohta 8.

<sup>244</sup> Esimerkiksi Bennet Moses puhuu teknologiasääntelyn tutkimuksen yhteydessä sosioteknisistä käytännöistä tutkimuksen perspektiiviä laajentavana käsitteenä. Käsitteen avulla voi siten nähdäkseni laajentaa myös sääntelyn perspektiiviä.

<sup>245</sup> Ks. Balwin et al. 2012 s. 2-3 & Black 2002 s. 2. Julie huomauttaa, että englannin kielen sanalle "regulation" ei ole useinkaan sisällöltään samaa käännöstä englanninkielisten maiden ulkopuolella. Nähdäkseni näin on suomen kielen osalta. Käyttämäni käännös sääntely ei täysin kuvaa sanan englannin kielen merkityssisältöä.



Teknologian käsite typistyykin usein viimeaikaisen kehityksen synnyttämiin uusiin teknologioihin.<sup>246</sup> Bennet Moses antaa tästä oivan esimerkin todetessaan, että tieliikennesääntelystä puhutaan enää harvoin teknologiasääntelyn kontekstissa.<sup>247</sup> Poikkeuksiakin on, kuten vaikka Koops, joka on määritellyt teknologian olevan *“laaja joukko erilaisia välineitä ja menetelmiä, joita ihmiset käyttävät muuttaakseen tai sopeutuakseen ympäristöönsä”*<sup>248</sup>. Nähdäkseni tämän määritelmän avulla teknologiasta on mahdollista keskustella ilman uutuuteen liitettävää historiattomuutta ja näin ollen arvioimaan tarkemmin kunkin teknologian erityispiirteitä ja sääntelyssä käytettävän keinovalikoiman tarpeellisuutta ja sopivuutta.<sup>249</sup>

Tekoälysäädöksen johdanto-osassa tekoälyn todetaan olevan nopeasti kehittyvä joukko teknologioita, jolla edistetään taloudellista, sosiaalista ja ympäristön hyvinvointia useilla yhteiskunnan osa-alueilla.<sup>250</sup> Kuten aineiston analyysiä koskevassa jaksossa esitettiin, tekoälyjärjestelmän määritellään puolestaan konepohjaiseksi järjestelmäksi, joka on suunniteltu toimimaan vaihtelevilla autonomian asteilla ja voivan muuttua käyttöönoton jälkeenkin, tuottaen ennusteita, sisältöä, suosituksia tai päätöksiä, jotka vaikuttavat ympäristöön, jonka kanssa järjestelmä on vuorovaikutuksessa joko fyysisessä tai digitaalisessa ulottuvuudesta.<sup>251</sup> Siten leimallista järjestelmille on tavallisen tietojenkäsittelyn sijaan oppiminen, mallintaminen ja päättely.

Tekoälyn määritelmää ei toisin sanoen ole sidottu tiettyyn teknologiaan, mutta tekoälyjärjestelmän määritelmän mukaisesti tietyt teknologian ominaisuudet altistavat järjestelmän sääntelylle. Tekoälysäädöksessä ei kiinnitetä paljoakaan huomiota eri teknologioiden erityispiirteiden vertailulle tai tuoda esiin kehitystä automaattisista järjestelmistä autonomisiin. Tältä osin tekoälysäädöksen perusteella on hankala luoda käsitystä historiallisesta kehityskaaresta ja arvioida tekoälyjärjestelmien tunnistettuja erityispiirteistä suhteessa muihin järjestelmiin.

---

246 Bennet Moses 2013 s. 4. Esimerkiksi Brownsword 2008 s. 1 listaa kirjassaan laajan joukon erilaisia teknologioita. Näitä ovat muun muassa nano-, telekommunikaatio-, viestintä- ja neurotiedeteknologiat.

247 Bennet Moses 2013 s. 4-5.

<sup>248</sup> Koops 2010 s. 311.

249 Esimerkiksi Hakkarainen on todennut, että näennäisesti uusia aiheita tutkittaessa on joskus hankala sijoittaa tutkimusaihetta osaksi laajempaa historiallista jatkumoa. 2022 s. 6.

250 Tekoälysäädöksen johdanto-osan perustelukappale 4.

251 Tekoälysäädös artikla 3(1).

Siten myös keinovalikoiman sopivuuden arviointi kattavasti on pelkän tekoälysäädöksen avulla vaikeaa. Jos tarkastelemme keinovalikoimaa puhtaasti tekoälysäädöksen kontekstissa, valitut lähestymistavat vaikuttavat intuitiivisesti järkeviltä. Esimerkiksi läpinäkymättömyyden erikoispiirteeseen vastataan tiedonantovaatimuksilla. Tiettyjen vaatimusten, kuten ihmisvalvonnan osalta voimme kuitenkin todeta, että tekoälysäädöksen ja autonomisten järjestelmien kontekstista irtaantuminen tuo esiin seikat, jotka osoittavat keinovalikoiman tehottomuutta.

#### 4.2.3 Sääntelyn määritelmä

Nimenomaan sääntelyn (eng. “regulation”) käsitteen problematisoinnit ovat olleet teknologiakontekstissa suosittuja.<sup>252</sup> Valtioiden tarkoituksella asettaman lainsäädännön lisäksi sääntelyllä voidaan viitata kaikenlaiseen yhteiskunnalliseen ja taloudelliseen vaikuttamiseen.<sup>253</sup> Laajassa mielessä sääntelyksi voidaan ymmärtää myös tällaisten vaikuttamiskeinojen tahattomat vaikutukset.<sup>254</sup> Laajasti ymmärretyn sääntelyn käsitteen käytöllä on omat etunsa verrattuna pelkästään lainsäädännöstä (eng. “law”) puhumiseen: ensinnäkin se mahdollistaa muiden yhteiskunnallisten (tahattomien) vaikutusten huomioimisen kuten markkinoiden vaikutukset; toiseksi se korostaa muun kuin horisontaalisen sääntelyn mahdollisuutta, esimerkiksi eri toimialojen itse luomat säännöt; kolmanneksi teknologioiden laajassa viitekehyksessä on luotu paljon standardeja ja soft law:ta, jotka on mahdollista huomioida.<sup>255</sup> Siten laajasti ymmärretty sääntelyn määritelmä auttaa paremmin teknologian moninaisten vaikutusten hahmottamisessa, kuin pelkästään lainsäädäntöön fokuoituminen.<sup>256</sup>

Kuten HCI-jaksosta huomasimme, ihmiskeskeisten tekoälyjärjestelmien kehittämiseksi on luotu kansainvälisiä standardeja.<sup>257</sup> Tekoälysäädöksen trilogineuvotteluiden päätteeksi laaditussa säädösversiossa sovinnon tiivistelmää

---

252 Bennet Moses 2013 s. 4-9. Bennet Moses huomauttaa, että toisin kuin teknologian käsitteen yhteydessä tämä ei aiheuta useinkaan ongelmia, sillä sääntelyn käsitteen suhteen ollaan määritelmässä eksakteja.

253 Baldwin et al. 2011 s. 309-310.

254 Black 2002 s. 1-2. Vrt. esim. Koops, jonka mukaan sääntely viittaa vain tahalliseen vaikuttamiseen (“the intentional influencing of someone’s or something’s behaviour”) 2010 s. 309.

255 Bennet Moses 2013 s. 3-4.

256 Ibid.

257 IEEE 2024 Autonomous Systems Industry Connections.

koskevat kohdat paljastavat, että tekoölyjärjestelmän määritelmän muuttamiseen vaikutti yhdenmukaistamisen tarve kansainvälisten järjestöjen määritelmien kanssa. Erityisesti OECD:n määritelmän kanssa.<sup>258</sup> Mielenkiintoisesti komission ehdottomassa luonnoksessa tekoölysäädökseksi oli myös maininta, jonka mukaan tekoölysäädökseen ehdotetut vaatimukset muodostavat osan alalla jo noudatetuista parhaista käytännöistä.<sup>259</sup> Alussa esiteltyjen valmisteluasiakirjojen ja itse tekoölysäädöksen luotettavan tekoölyn vaatimusten välillä voimme puolestaan todeta yhteyden ja siten asiantuntijatyöryhmien valmistelemilla näkemyksillä on suora vaikutus lainsäädäntöön. Siten soft law:lla ja standardeilla on ollut muovaava vaikutus tekoölysäädökseen ja alan parhaiden käytäntöjen kodifointi osoittaa, että säädös ei ole täysin horisontaalisesti luotu. Säädöksen tavoitteiden mukaisesti se kuitenkin harmonisoi käytäntöjä ja siten myös oikeudellinen vuorovaikutuskäsitys on omiaan harmonisoitumaan.

Tutkielman STS-viitekehityksessä teknologia mielletään myös sosiaalisen sääntelyn välineeksi. Ainakin, jos sääntelyksi lasketaan myös tahattomat vaikutukset. Tekoölysäädöksessä on tunnistettu mahdollisuus välineellistää tekoölyteknologia sosiaalisen sääntelyn käyttöön. Tämä käy ilmi erityisesti kielletyistä tekoölyteknologioiden käyttötavoista, joissa tarkoituksena on kaventaa ihmisten vapaata toimijuutta. Tekoölysäädöksessä kielletään yksiselitteisesti sellaiset tekoölyjärjestelmät, jotka kykenisivät hyväksikäyttämään ihmisen havaintokykyä manipulatiivisella tavalla eli subliminaaliset tekniikat. Järjestelmät eivät saa haitata ihmisen vapaata päätöksentekoa.<sup>260</sup> Tosin sanoen vuorovaikutus, joka vie ihmiseltä toimijuuden suhteessa tekoölyjärjestelmään on kiellettyä. Kieltoon liittyy nimenomaan manipulointi, jolla houkutellaan luonnollinen henkilö ei haluttuun tai toivottuun toimintaan.<sup>261</sup> Kielto ilmentää ihmistoimijuuden kunnioittamisen vaatimusta luotettavan tekoölyn edellytyksenä.<sup>262</sup>

---

<sup>258</sup> Euroopan komissio 2024 Interinstitutional File: 2021/0106(COD) s. 3.

<sup>259</sup> Tekoölysäädös on siten myös esimerkki vertikaalisten käytäntöjen ja standardien merkityksestä lainsäädännölle. Ks. Euroopan komissio (2021) 206 final 2021/0106(COD), Ehdotukseen sisältyvien säännösten yksityiskohtaiset määritelmät, 5.2.3.

<sup>260</sup> Tekoölysäädöksen johdanto-osan perustelukappale 29 & artikla 5.

<sup>261</sup> Ibid.

<sup>262</sup> Samalla kohta osoittaa ihmistoimijan haavoittuvuuden vuorovaikutuksessa tekoölyjärjestelmään, mikäli järjestelmää käytetään välineenä sosiaaliselle sääntelylle. Kohta paljastaa, että ihmisen havainnointi- ja tietojenkäsittelykykyjen heikkouksia voidaan vuorovaikutuksessa käyttää hyväksi toimijuutta loukkaavalla tavalla.

Lisäksi reaaliaikaisten biometriseen tunnistukseen nojaavien järjestelmien käytön julkisilla paikoilla todetaan olevan erittäin pitkälle menevää perusoikeuksiin puuttumista. Järjestelmien käytöllä saattaa olla vaikutuksia suuren väestömäärän yksityisyyteen, se saattaa herättää tunteen jatkuvasta tarkkailusta ja epäsuorasti kaventaa järjestäytymisvapauden sekä muiden perusoikeuksien toteutumista.<sup>263</sup> Kohta on mielenkiintoinen, sillä siinä esitetään tietoisuuden järjestelmien käytöstä muuttavan ihmisten käytöstä nimenomaan toimijuutta kaventavalla tavalla. Ihmisen ja järjestelmän välillä ei edellytetä olevan suoranaista vuorovaikutusta vaan järjestelmän läsnäolo riittää aikaan saamaan vaikutuksen. Samoin tekoälyjärjestelmä, jonka tarkoituksena on luoda kasvojen tunnistukseen tähtääviä tietokantoja voi lisätä kokemusta massavalvonnasta ja on siten kiellettyä perusoikeuksia liikaa rajoittavana.<sup>264</sup> Nämä esimerkit ilmentävät mahdollisuutta tahattomaan sosiaalisen todellisuuden sääntelyyn teknologiaa hyödyntämällä.

#### 4.2.4 Yleisen edun perustelut teknologian käyttöönotossa

Teknologiasääntelyn tutkimuksessa on osoitettu, että usein teknologioiden käyttöönottoa ja kehitystä perustellaan yleisellä hyvällä.<sup>265</sup> Tätä on arvioitu kriittisesti toteamalla, että tieteellisen edistyneisyyden ja sitä kautta saavutettavilla hyödyillä annetaan automaattinen ja kyseenalaistamaton hyväksyntä uusien teknologioiden käyttöönotolle. Hyötyjen avulla siten legitimoidaan ja tarvittaessa perustellaan sääntelyn puutetta tai sääntelyhaluttomuutta.<sup>266</sup>

Kuten jo valmisteluasiakirjoista ilmeni, tekoälysäädöksessä on omaksuttu ajatus tekoälyteknologian hyödyistä koko yhteiskunnalle. Nämä vaikutukset syntyvät, koska tekoäly muun muassa parantaa ennakoitavuutta, optimoi toimintoja ja funktioalokaatiota sekä mahdollistaa personoitujen digitaalisten ratkaisujen luomisen yksilöille ja organisaatioille.<sup>267</sup> Uusien tekoälyteknologioiden mainitaan olevan olennaisia esimerkiksi ympäristökriisin ratkaisemiselle ja siten tuottavan merkittäviä hyötyjä.<sup>268</sup> Tekoälysäädöksessä hyötyajattelulla ei kuitenkaan oikeuteta

---

<sup>263</sup> Tekoälysäädöksen johdanto-osan perustelukappale 32.

<sup>264</sup> Tekoälysäädös artikla 5(1(e)).

<sup>265</sup> Ks. esim. Brownsword 2017 luku 5.

<sup>266</sup> Ibid.

<sup>267</sup> Tekoälysäädöksen johdanto-osan perustelukappale 4.

<sup>268</sup> Ibid. Kappaleessa mainitaan esimerkkisektoreina terveydenhuolto, maatalous, elintarviketeollisuus, koulutus, media, urheilu, kulttuuri, infrastruktuuri, energia, energiatehokkuus,

sääntelemättömyyttä, sillä samanaikaisesti tunnistetaan, että tekoäly voi sen erityiseen sovellukseen ja käyttöön liittyvistä olosuhteista riippuen aiheuttaa riskejä ja vahingoittaa unionin lainsäädännöllä suojattuja yleisiä etuja ja oikeuksia.<sup>269</sup> Tämä luo sääntelytarpeen.

Vaikka tekoälysäädöksen tasapainottamista vaativat päämäärät eli teknologian kehitys ja sen aiheuttamat riskit ilmenevät usein, yleisen edun ensisijaisuutta on silti havaittavissa. Yleinen etu saattaa vaatia kiellettyjen menettelyiden sallimista ja tekoälyjärjestelmäkäytäntöjen muille perusoikeuksille muodostaman riskin tai rajoittamisen sivuuttamisen. Esimerkiksi artikla 54 mahdollistaa testiympäristössä tekoälyjärjestelmien opettamisen muihin käyttötarkoituksiin kerätyillä henkilötiedoilla, kun tekoälyjärjestelmä kehitetään merkittävän yleisen edun turvaamiseksi. Myös esimerkiksi vaatimustenmukaisuusarviointi voidaan sivuuttaa, jos yleinen turvallisuus asianmukaisesti perustellussa kiireellisessä tilanteessa edellyttää korkean riskin tekoälyjärjestelmän käyttöönottoa ilman arviota.<sup>270</sup>

Verrattuna eurooppalaiseen lähestymistapaan Yhdysvaltojen tekoälystrategiassa on nähtävillä huomattavasti vahvempi taloudellinen painotus, jossa ei keskitytä eurooppalaiseen tapaan riskinäkökulmiin. Tämä on luettavissa jo pelkästään strategiaa edeltäneiden ja sitä seuranneiden politiikkadokumenttien nimistä.<sup>271</sup> Talouden lisäksi tekoälyn tutkimuksen ja kehityksen keskiössä Yhdysvalloissa on tulevaisuuden työelämän muutokset ja niihin varautuminen sekä tekoälyn sotilaalliset sovellutukset.<sup>272</sup> Toisaalta sama vertailututkimus osoittaa maailman neljänkymmenen eri tekoälystrategian fokuoivan tekoälyn yleishyödyllistä ulottuvuutta mutta myös korostavan ihmiskeskeistä suunnittelua.<sup>273</sup>

Yleishyödyllisen näkökulman korostaminen ei automaattisesti tarkoita sääntelystä pidättäytymistä. Tekoälysäädöksen punninta hyötyjen ja haittojen välillä siis legitimoii sääntelyn tarpeen. Säädos on siten kahden selkeän päämääränsä osalta jännitteinen. Tekoälyjärjestelmät mahdollistavat hyvinvointia lisäävät innovaatiot

---

liikenne, logistiikka, julkishallinto, turvallisuus, oikeuslaitos ja ympäristövalvonta. Mahdollisuudet tunnistetaan siten lähes kaikilla ajateltavilla sektoreilla.

<sup>269</sup> Tekoälysäädöksen johdanto-osan perustelukappale 5.

<sup>270</sup> Tekoälysäädöksen Artikla 47(1a).

<sup>271</sup> Ks. Shackelford et al. 2021 s. 38-40.

<sup>272</sup> Ibid.

<sup>273</sup> Ibid. s. 87.

mutta toisaalta muodostavat merkittävän riskin perusoikeuksille. Tekoälyjärjestelmien kehitykseen ja käyttöön liitetyt riskit luovat tarpeen toiselle päätavoitteelle eli terveyden, turvallisuuden ja perusoikeuksien suojelulle. Ilman tekoälyjärjestelmiin liitettyä uhkaa, ei olisi tarvetta säädöksessä omaksutulle sääntelylle. Koulu onkin nostanut esiin, että tekoälyn etiikkaa koskevat ohjeet ja standardit joutuvat väistämättä maalaamaan ensin uhkakuvaa, jotta tarve standardeille syntyy. Standardit luovat ratkaisun ongelmaan. Koulu huomauttaa myös, että monesti soft law-instrumentit ovat pohjana sitovan lainsäädännön luomisessa.<sup>274</sup> Näin on käynyt myös tekoälysäädöksessä. Vaikuttaa selvältä, että tekoälysäädös perustuu samalle asetelmalle uhkien esittämisestä ja ratkaisujen tarjoamisesta. Lisäksi omaksutut uhkakuvat ja ratkaisut niiden taklaamiseksi ovat linjassa politiikkadokumenteissa esitetyn kanssa.

#### 4.2.5 Teknologian kaksoisrooli sääntelyssä

Teknologiasääntelyssä joudutaan siis väistämättä punnitsemaan teknologian mahdollisuuksien ja haasteiden suhdetta, jotta legitiimin yhteiskunnan kulmakiviä ei uhrata tehokkuudelle tai muille teknologian hyödyille. Brownsword on huomauttanut, että sopivan tasapainon löytäminen teknologioiden edistämisen ja rajoittamisen välillä johtaa usein perinteisten oikeudellisten menettelyiden täydentämiseen tai korvaamiseen teknologioilla.<sup>275</sup> Yksi teknologiasääntelyn oikeudelle asettamista haasteista johtuukin teknologian kaksinaisesta roolista sääntelyssä. Se on yhtäältä sääntelyn kohde (eng. “regulatory target”), että sääntelyn väline (eng. “regulatory tool”).<sup>276</sup> Kun teknologia on sääntelyn kohteena, keskitytään hallitsemaan teknologian haittoja ja riskejä. Kun teknologiaa käytetään sääntelyn välineenä, kyseessä voi olla esimerkiksi valvontakameroiden käyttö.<sup>277</sup> Kuten jo oikeuden digitalisaatiota koskevassa kohdassa todettiin teknologian vaikutukset yltyvät kaikkiin oikeudellisiin käytäntöihin.

Kuvaava esimerkki kehityksestä tekoälysäädöksen ulkopuolelta on sääntelyteknologiaksi (eng. “RegTech”) nimitetty ohjelmisto- ja infrastruktuuriratkaisujen ala, jossa pääasiallisena tavoitteena on luoda

---

<sup>274</sup> Koulu 2020 s. 14.

<sup>275</sup> Brownsword 2008 s. 2.

<sup>276</sup> Ibid.

<sup>277</sup> Bennet Moses 2013 s. 3. Tai esimerkiksi jonkinlainen uusi sääntelyteknologian innovaatio.

teknologiaratkaisuja yrityksille niiden toimeenpannessa teknologiasäätelyä samalla säilyttäen innovatiivisuutensa ja kilpailukykyä.<sup>278</sup> Teknologian hyödyntäminen onkin laaja-alaista eritoten lainsäädännön toimeenpanon ja valvonnan apuvälineenä. Esimerkiksi tekoälysäädöksessä korostetaan, että lainvalvontaviranomaisten kuten tullin ja verottajan käyttämät tekoälyjärjestelmät, joilla valvotaan unionin rahanpesulainsäädännön noudattamista ei muodosta korkean riskin järjestelmää.<sup>279</sup>

#### 4.2.6 Teknologiasäätelyn tehokkuus ja kohdentuminen

Muita teknologiasäätelyn haasteita on säätelyn tehokkuuteen, kohdentumiseen (eng. “connection”), legitimiin ja sääntöjen noudattamiseen liittyvät ongelmat.<sup>280</sup> Kohdentumisen ongelma muodostuu, kun lainsäädäntö ei vastaa teknologian kehitystasetta. Kohdentumista tai sen katkeamista voidaan havainnollistaa erottelemalla lainsäädännössä ja säätelyssä olevat katkokset jakson alussa esiteltyyn laajasti ymmärretyn säätelyn käsitteen avulla. Olemassa oleva lainsäädäntö kieltää esimerkiksi tekijänoikeudella suojatun digitaalisen sisällön jakamisen ilman lupaa. Siten lainsäädännön kohdentumisessa ei ole katkoa. Sisällön jakamisen helppous on kuitenkin vaikuttanut yhteiskunnallisiin käytäntöihin niin, että jakamismäärät ovat lisääntyneet lainsäädännöstä huolimatta, ja siten käsillä on säätelykatko.<sup>281</sup> Kuten tekoälysäädöksen valmisteluasiakirjoissa todettiin, tekoälyn kehittäjiin ja käyttöönottajiin sovelletaan jo nyt perusoikeuksia, esimerkiksi yksityisyyttä ja syrjimättömyyttä. Kuitenkin tekoälyn erityispiirteiden, kuten läpinäkymättömyyden vuoksi, nykyisen lainsäädännön soveltaminen ja toimeenpano voi olla haastavaa ja näin ollen syntyy tarve uudelle lainsäädännölle.<sup>282</sup> Brownswordin mukaan tällä tarkoitetaan lainsäädännön uudelleen kohdentamista (eng. “reconnection”).<sup>283</sup>

Vaikuttaa siltä, että tekoälyteknologioiden osalta on havaittu sekä säätelykatkon olemassaolo että näkemys lainsäädännön uudelleen kohdentamisen tarpeesta. Esimerkiksi syrjimättömyyden, monimuotoisuuden ja oikeudenmukaisuuden periaatteiden mukaisesti tekoälyjärjestelmiä tulee kehittää ja käyttää tavalla, joka koostuu erilaisista toimijoista ja edistää kaikkien yhtäläisiä

---

<sup>278</sup> Fehlinger 2023 s. 7-8.

<sup>279</sup> Tekoälysäädöksen johdanto-osan perustelukappale 59.

<sup>280</sup> Brownsword 2008 s. 2.

<sup>281</sup> Bennet Moses 2013 s. 6.

<sup>282</sup> Euroopan komissio COM(2020) 65 final s. 11.

<sup>283</sup> Brownsword 2008 s. 160-184.

osallistumismahdollisuuksia sekä vähentää syrjintää.<sup>284</sup> Syrjimättömyyttä korostetaan erityisesti lainvalvonnan ja -käytön konteksteissa. Johdanto-osan perustelukappale 42 puolestaan käsittelee syyttömyysolettamaa ja tekoälyjärjestelmän käyttöä ennakoitaessa luonnollisen henkilön mahdollisuutta syyllistyä rikokseen. Syyttömyysolettaman mukaisesti luonnolliset henkilöt tulee tuomita vain varsinaisen käytöksensä perusteella. Kohdan mukaan luonnollisten henkilöiden tuomitseminen tai rikoksentekoalttiuden arviointi pelkästään tekoälyjärjestelmän ennustaman käytöksen perusteella perustuen esimerkiksi heidän profilointiin, luonteenpiirteisiin tai ominaisuuksiin kuten kansallisuuteen, syntymä- tai asuinpaikkaan on kiellettyä. Tuomitseminen voi tapahtua vain objektiivisesti varmistettujen faktojen valossa ja luonnollisen henkilön toimesta. Kohdan tarkoituksena vaikuttaa olevan kielto kiertää syyttömyysolettamaa käyttämällä tekoälyjärjestelmää. Voimassa olevan lainsäädännön vuoksi erillistä kieltoa ei tarvittaisi, mutta se ilmentää halua korostaa ihmistoimijoiden tasapuolista kohtelua myös tekoälyjärjestelmiä käytettäessä.

Samoin tekoälysäädöksessä vaaditaan, että järjestelmien käyttö työvälineenä viranomaisissa tai lainkäytössä ei saa johtaa epätasa-arvoon tai poissulkemiseen. Puolustus oikeuden käyttämiseen tulee säilyä mahdollisuus, ottaen huomioon erityisesti vaikeudet tarkoituksenmukaiseen tiedonsaantiin tekoälyjärjestelmän toiminnoista ja siitä johtuvaan vaikeuteen haastaa tuloksia oikeudessa.<sup>285</sup> Jälleen kiellot kohdistuvat jo sinänsä selkeästi kiellettyyn menettelyyn eli tässä tapauksessa syrjintään viranomaisessa tai tuomioistuimessa tai vahvistetaan oikeus positiivisen oikeuden käyttöön. Mielestäni molemmat syrjintään liittyvät esimerkit osoittavat tekoälysäädöksen otaksuvan tekoälyjärjestelmien muuttavan vuorovaikutussuhteita siinä määrin, että aiemman sääntelyn ei katsota analogioiden avulla olevan riittävää, vaan kiellot tulee toistaa uudessa muuttuneessa kontekstissa. Lainsäädäntö tulee uudelleen kohdentaa tekoälyteknologioihin, sillä tekoälyjärjestelmien erityispiirteiden vuoksi vanha normisto ei ole ongelmitta sovellettavissa.

Valmisteluasiakirjoissa lainsäädännössä tunnistettiin suoranainen katko ihmisen valvonnan ja läpinäkyvyyden osalta. Tekoälysäädöksen valmisteludokumenteissa todettiin, että ihmisvalvonnan ja läpinäkyvyyden vaatimuksia ei käsitellä olemassa

---

<sup>284</sup> Tekoälysäädös johdanto-osan perustelukappale 27.

<sup>285</sup> Tekoälysäädös johdanto-osan perustelukappale 59.



olevassa sääntelykehyksessä ja siten kyseiset vaatimukset tulee sisällyttää tulevaan lainsäädäntökehykseen.<sup>286</sup> Läpinäkyvyys ilmenee erityisesti siten, että ihmisen tulee kyetä tekemään tekoälyjärjestelmää koskevia informoituja päätöksiä.<sup>287</sup> Tietoisuus tekoälyjärjestelmille altistumisesta ja oikeus selitykseen on läpinäkyvyyden toinen ulottuvuus. Ihmisen luoman sisällön katsotaan tekoälysäädöksessä olevan aitoa verrattuna tekoälyjärjestelmien luomaan synteettiseen sisältöön.<sup>288</sup> Erilaiset tekoälyjärjestelmät pystyvät luomaan massoittain synteettistä sisältöä, jonka erottaminen ihmisen luomasta ja autenttisesta sisällöstä voi olla ihmiselle lähestulkoon mahdotonta. Synteettisen ja epäaidon sisällön katsotaan muodostavan riskin tietökosysteemien yhtenäisyydelle ja luotettavuudelle. Tämä luo kasvavaa alustaa misinformaatiolle ja manipulaation muodoille erityisesti petosten, identiteettivarkauksien ja kuluttajien harhaanjohtamisen osalta. Toisin sanoen tekoälyjärjestelmien luoma synteettinen sisältö on uusi riski, jonka hallitsemiseksi tarvitaan uutta lainsäädäntöä.

Edellä esitettyjen riskien vuoksi tekoälyjärjestelmien tarjoajien tulee luoda tekniset ratkaisut, jotka mahdollistavat järjestelmän luoman sisällön havaitsemisen ja merkitsemisen ja näin ollen tekevät mahdolliseksi erottaa ihmisen luoman sisällön.<sup>289</sup> Sama läpinäkyvyyden vaatimus halutaan ulottaa tekoälyjärjestelmän käyttöönottajalle, jonka tulee selvästi ja erotettavasti ilmoittaa, että sisältö on tekoälyn luomaa tai manipuloimaa.<sup>290</sup> Vaatimukset palautuvat läpinäkyvyyden kautta ihmistoimijuuden arvostamiseen. Toisaalta se ilmaisee, että vain ihmistoimijan luoma sisältö on autenttista ja korostaa näin ihmisen arvokkuutta suhteessa järjestelmiin. Lisäksi se viittaa siihen, että järjestelmän luoma sisältö on ihmisen luomaan sisältöön nähden riskialttiimpaa, jonka vuoksi ihmisen on oltava siitä

---

286 Euroopan komissio COM(2020) 65 final s. 14-15.

287 Vaatimukseen kuuluu muun muassa kehitysvaiheessa tapahtuva oikea teknisten elementtien soveltaminen, toimenpiteet käytön aikana, oikeaoppinen tulosten tulkinta, ja mikäli tekoälyn käytöllä on vaikutuksia luonnollisiin henkilöihin, asianmukainen arviointi heihin kohdistuvista vaikutuksista. Ks. tekoälysäädöksen johdanto-osan perustelukappale 20. Ihmisiltä odotettava AI-lukutaito on käsitteellistys tekoälyjärjestelmien ymmärrykselle. AI-lukutaidolla viitataan määritelmien mukaan tarjoajan, käyttäjän ja kohteena olevan luonnollisen henkilön taitoihin, tietoihin ja ymmärrykseen, joka mahdollistaa säädöksen mukaisten oikeuksien ja velvollisuuksien rajoissa tekoälyjärjestelmän valistuneen käytön. Ks. Tekoälysäädöksen artikla 3(56).

288 Tekoälysäädöksen johdanto-osan perustelukappale 133.

289 Ibid.

290 Ibid.

tietoinen. Vuorovaikutuksen näkökulmasta pyrkimys on nähdäkseni hyvä, sillä se lisää avoimuutta ihmisen ja järjestelmän välillä.

Yhtä lailla vuorovaikutuksen läpinäkyvyyttä ja avoimuutta ilmentää korkean riskin tekoälyjärjestelmien käyttöönottajille asetettu vaatimus tiedottaa tekoälyjärjestelmien vaikutuksille ja käytölle altistuvia henkilöitä. Kun käyttöönottajat käyttävät korkean riskin järjestelmiä luonnollisia henkilöitä koskevassa päätöksenteossa, käyttö pitää saattaa luonnollisten henkilöiden tietoon. Tiedon pitää sisältää järjestelmän aiottu käyttötarkoitus ja päätöksen tyyppi. Lisäksi oikeudesta selitykseen on tiedotettava.<sup>291</sup> Oikeus selitykseen on luonnollisen henkilön oikeussuojakeino. Tietyissä korkean riskin toiminnoissa ihmisen tulee saada selkeä ja tarkoituksenmukainen (eng. “meaningful”) selitys, joka mahdollistaa luonnollisen henkilön oikeuksien käyttämisen.<sup>292</sup> Velvollisuuksilla korjataan käsillä ollut sääntelykatko ja pyritään mahdollistamaan ihmisille oikeuksiensa valvominen ja siten vahvistamaan ihmisen toimijuutta.

#### 4.2.7 Sääntelyvalintojen vaikutuksesta

Jo mainittujen haasteiden lisäksi monet tutkijat nostaneet haasteeksi sääntelyn oikea-aikaisuuden. Onnistunut sääntely vaatii muiden ohella momentumin. Tätä nimitetään Collingridgen dilemmaksi. Jonkin teknologian kehittyessä sen varhainen sääntely on haastavaa, sillä vaikutuksia ei vielä tiedetä. Sen sijaan liian myöhään tapahtuva puuttuminen voi olla ongelma, koska teknologia juurtuu ja muutosten tekeminen on kallista ja työlästä.<sup>293</sup> Sääntelyn kohdalla voidaan havaita siten sama vaikutus kuin HCI-tutkimuksessa juurtuneen suunnitteluteorian kohdalla. Tämä huomio vie takaisin teknologiasääntelyn tutkimuksen haasteisiin. Sääntelyn onnistumisen todennäköisyys kasvaa, jos sääntelyn pohjalla on perusteellista tutkimustietoa. Toki sääntelyyn liittyy tästäkin huolimatta epävarmuuksia ja arvaamattomia vaikutuksia.<sup>294</sup>

---

291 Tekoälysäädöksen johdanto-osan perustelukappale 93.

292 Tekoälysäädöksen johdanto-osan perustelukappale 171.

293 Bennet Moses 2013 s. 7.

294 Sankari et al. 2022 s. 189.

## 5 Oikeudellistuvan vuorovaikutuskäsityksen ydin ja ongelmat

### 5.1.1 Ihmisvalvonta luotettavan tekoälyn edellytyksenä

Ihmisvalvonnan vaatimuksen taustalla on ihmistoimijuuden säilyttäminen ja järjestelmän toimijuuden palauttaminen ihmistoimijaan. Ihmisvalvonnan vaatimuksella tarkoitetaan yksinkertaisesti sitä, että ihmisen täytyy asianmukaisesti pystyä kontrolloimaan ja valvomaan tekoälyjärjestelmiä.<sup>295</sup> Ihmisvalvonnan vaatimus omaksuttiin luotettavan tekoälyn edellytykseksi jo tekoälysäädöstä koskevissa politiikkadokumenteissa.<sup>296</sup> Ihmisvalvontaa koskeva artikla 14 ja sitä käsittelevät johdanto-osan kappaleet ovat tekoälysäädöstä leimaavan hierarkkisen vuorovaikutuksen hahmotustavan selkeimmät ilmentymät arvottavien tehtävien kuvausten lisäksi. Kohdat nostavat ihmisen järjestelmien yläpuolelle valvovaan ja kontrolloivaan rooliin ja tilaa vuorovaikutuksen joustavuudelle ei säädöksessä anneta. Kuten analyysisiosiossa totesin, tämä on yksi vahvimmin dikotomiseen vuorovaikutuskäsitykseen viittaavista vuorovaikutusmuuttujista.

Tekoälysäädökseen omaksuttu ihmisvalvonnan toteutusperiaate on niin sanottu HIC-periaate (eng. "Human in Command"), joka tarkoittaa mahdollisuutta valvoa järjestelmän yleistä toimintaa.<sup>297</sup> Tekoälysäädöksessä todetaan, että ihmisvalvonnan vaatimus käsittää koko järjestelmän elinkaaren. Ihmisen kuuluu varmistua siitä, että käyttää systeemiä sen käyttötarkoituksen mukaan. Erityisesti ihmisvalvonnan vaatimus edellyttää, että järjestelmät sisältävät käyttörajoituksia, jotka tottelevat ihmistoimijaa ja joita järjestelmä ei voi itse ohittaa. Järjestelmän tulee ohjeistaa ja tukea ihmistä hänen arvioidessa, milloin ihmisen suorittama interventio on tarpeellinen riskien tai negatiivisten seurausten välttämiseksi. Tähän sisältyy tarvittaessa järjestelmän pysäyttäminen ja päätös olla käyttämättä sen tuotosta.<sup>298</sup>

Tekoälysäädöksen politiikkadokumentteja on tutkittu ihmisvalvonnan taustaoletusten selvittämiseksi. Tutkimuksessa on huomautettu, että poliittiset päätökset voivat muuntua oikeudellisiksi käsitteiksi, jotka operationalisoidaan osaksi oikeusjärjestelmää. Mikäli päätösten taustalla olevat oletukset eivät ole paikkaansa pitäviä, on riskinä sääntelyn tarjoaman suojan tehottomuus tai sen tavoitteiden

<sup>295</sup> Tekoälysäädöksen johdanto-osan perustelukappale 27.

<sup>296</sup> Euroopan parlamentti 2019 luku I: Luotettavan tekoälyn perusta s. 13-16.

<sup>297</sup> Euroopan parlamentti 2019 s. 19.

<sup>298</sup> Tekoälysäädöksen johdanto-osan perustelukappale 73 & artikla 14.

saavuttamisen mahdottomuus.<sup>299</sup> Useimmiten tällaisia taustaoletuksia ei kuitenkaan kirjoiteta auki politiikkadokumentteihin. Näin ollen ne pysyvät keskustelun ulkopuolella.<sup>300</sup> Poliitiikkadokumenteista omaksuttu ihmisvalvonta on nyt operationalisoitu oikeusjärjestelmän käsitteeksi.<sup>301</sup> Kuten HCI-osiossa ja aineiston analyysissä osoitin, se on todettu keinona ilmeisen tehottomaksi. Tekoälysäädöksessä oikeudellistuvan ihmisen ja järjestelmän vuorovaikutuksessa se on kuitenkin yksi keino varmistaa oikeusvarmuus ja perusoikeuksien suojele. Näin ollen voimme kyseenalaistaa sääntelyn tarjoaman suojan tehokkuuden ihmisen valvonnan osalta. Myöskään tekoälysäädöksessä ei käsitelty ihmisvalvonnan taustaoletuksia, siten politiikkadokumentteja koskevan tutkimuksen toteamalla tavalla oletamat ovat pysyneet oikeudellisen keskustelun ulkopuolella.

Vaikka politiikkadokumenteissa ja oikeuskirjallisuudessa korostetaan ihmisvalvonnan vaatimuksen uutuutta, nähdäkseni se ei ole täysin uusi edellytys. Esimerkiksi Enqvist esittää tekoälysäädöksen tuovan ihmisvalvonnan vaatimuksen uutena EU-lainsäädäntöön.<sup>302</sup> Mahdollisuus jonkin asteiseen valvontaan on jo aiemmin esitetty kontrollin vaatimuksen kautta, ja tätä on edellytetty vähintään implisiittisesti.<sup>303</sup> Unionin lainsäädännössä erityisesti liikennesektorilla on pitkä sääntelyhistoria ihminen-kone vuorovaikutuksesta.<sup>304</sup> Lainsäädännössä jo esitettyjä vaatimuksia on muun muassa se, että järjestelmien tulee välittää ihmisoperaattorille viestejä ja varoituksia, jotka vaativat välitöntä huomiota valvovalta ihmiseltä.<sup>305</sup> Kuitenkin kiistatta ihmisvalvonnan vaatimuksen painoarvo on nyt oikeudellistuvassa vuorovaikutuksessa omaa luokkaansa ja kyseisen käsitteen siirtyminen

---

<sup>299</sup> Koulu 2020 s. 15.

<sup>300</sup> Ibid. s. 18.

<sup>301</sup> Ks. esim. Enqvist 2023 s. 509.

<sup>302</sup> Ibid.

<sup>303</sup> Ks. esimerkiksi esteettömyysvaatimukset Euroopan parlamentin ja neuvoston direktiivissä (EU) 2019/882 liite 1 jakso 1.2.

<sup>304</sup> Käsitteenä ihminen-kone vuorovaikutus on ollut 2000-luvulta asti liikennesektorin sääntelyssä käytössä. Ks. Komission asetus (EY) N:o 2082/2000 3.2 määritelmät.

<sup>305</sup> Mielenkiintoisesti liikennesektorilla omaksuttu vuorovaikutuksen hahmotustapa vaikuttaa joustavammalta ja vähemmän tiukan hierakiselta. Järjestelmällä voidaan esimerkiksi nähdä olevan rooli ihmisen vireystilan valvojana ja sen on ryhdyttävä toimiin, jos epäaktiivisuutta havaitaan. Ks. Komission asetus (EU) N:o 1302/2014 kohta 4.2.9.3.1. Vastaava mahdollisuus järjestelmien suorittamaan ihmisen valvontaan onnettomuuksien ehkäisemiseksi jätettiin tekoälysäädökseen. Kyseisen kaltainen fyysisten tilojen tunnistaminen erotetaan tunteiden tunnistuksesta, joka muodostaa automaattisesti korkean riskin järjestelmän. Ks. Tekoälysäädöksen johdanto-osan perustelukappale 18.

lainsäädäntöön on osoitus siitä, että valvonta keinona vakiinnuttaa asemaansa ratkaisukeinona tekoälyjärjestelmien käytöstä aiheutuviin riskeihin.

Oikeuden on todettu olevan korostuneen kietoutunut käytettyyn kieleen ja käsitteisiin. Joitain oikeuden ilmiöitä ei ole välttämättä muuta kuin käsitteiden maailmassa.<sup>306</sup> Huomionarvoista kuitenkin on, että käsitteitä tai oikeudessa omaksuttuja mallinnuksia ei välttämättä luoda oikeudellisen kehityksen tai instrumenttien myötä. Niitä voidaan lainata myös muilta aloilta.<sup>307</sup> Mielenkiintoisesti ihmisvalvonnan vaatimus on otettu luotettavan tekoälyn edellytykseksi jo politiikkadokumenteissa. Ihmisvalvonnan vaatimuksen ensi kerran tekoälysäätelyn kontekstissa esittänyt korkean tason asiantuntijaryhmä ei koostu yksinomaan oikeudellisista asiantuntijoista. Pikemminkin ryhmässä on koostunut monialaisesti eri asiantuntijoista.<sup>308</sup> Pelkästään oikeuden antroposentrismi ei siis nähdäkseni selitä ihmisvalvonnan vaatimuksen omaksumista vaan vaikuttaa siltä, että ihmisen kontrollin säilyttämiselle on myös laajempi poliittinen halu. Tämä ei sinänsä ole yllättävää, sillä kuten toisessa jaksossa huomasimme, ihmisen kontrollin säilyttämiselle on HCI-traditiossakin ollut poliittisia vaikuttimia. Oikeudellisessa kontekstissa ymmärrettävästi tätä lähestymistapaa ei haluta kyseenalaistaa, sillä se mahdollistaa säätelyn ja vastuun kohdentamisen yhä ihmiseen. Siten oikeudellisesta näkökulmasta vältetään kysymyksenasetteluilta, jotka pakottaisivat miettimään vaihtoehtoja olemassa oleville oikeudellisille instituutioille.

Myös HCI-tutkimuksessa korostettiin, että koska ihmisen ja järjestelmien toimijuuden jakautuminen on esimerkiksi vastuukysymysten kannalta keskeinen seikka, vuorovaikutuksen ymmärrys ei ole vain järjestelmäsuunnittelijoiden asia. Perustason ymmärrys olisi olennaista säätelyn kehittäjille tarkoituksenmukaisen säätelyn varmistamiseksi.<sup>309</sup> Ottaen huomioon eettisten periaatteiden laatijoiden asiantuntijuus vaikuttaa lähes selvältä, että perustason ymmärrys ihmisten ja järjestelmien vuorovaikutuksesta on ollut säätelyn taustalla. Silti on huomattava, että kuilu poliittisten valmisteluasiakirjojen ja valmiin lainsäädännön välillä on pitkä.

---

<sup>306</sup> Ks. Koulu et al 2024: Solum 1992 s. 1232.

<sup>307</sup> Pyöreän pöydän keskustelu: Pöysti 20.9.2023.

<sup>308</sup> Euroopan parlamentti 2019 s. 50.

<sup>309</sup> Simmler & Frischknecht s. 239-240.

Koska valittuja keinoja ei ole jo todetulla tavalla perusteltu, on haastava arvioida, mihin olettamiin ihmisvalvonta luotettavan tekoälyn edellytyksenä perustuu.

Oikeudellisesta näkökulmasta ihmisvalvonta on laadunvarmistuskeino, jolla suojataan perusoikeuksia.<sup>310</sup> Tekoälysäädöksen 14 artikla on oikeussuojakeinona osin preventiivinen eli sillä pyritään ennaltaehkäisemään oikeuden loukkauksia ja siten lisäämään oikeusvarmuutta. Tämä poikkeaa hieman käytössä olevasta oikeussuojamekanismista, joka on omaksuttu tietosuojajakeinossa.<sup>311</sup> Esimerkiksi tietosuojajakeinossa oikeusturvakeinoksi automaattisten järjestelmien suorittaman päätöksenteon osalta on omaksuttu jälkikäteinen laadunvarmistus, jossa päätöksen kohteena olevalla henkilöllä on oikeus saada häntä koskeva päätös luonnollisen henkilön tarkistamaksi.<sup>312</sup> Ihmisosallisuuden säilyttämisen suosio oikeussuojamekanismina on siten säädösrajat ylittävää, vaikkakin eri toteutustapoja on. Tekoälysäädöksen myötä EU-lainsäädäntöön tulee toinen oikeussuojamekanismi ihmisestä etääntyneelle päätöksenteolle, joka on riippuvainen käytetystä teknologiasta. Kuitenkin myös tekoälysäädökseen on omaksuttu jälkikäteisiä suojamekanismeja. Esimerkiksi yllä kuvattu oikeus saada selkeä ja tarkoituksenmukainen selitys tekoälyjärjestelmän käytöstä edesauttaa yksilön mahdollisuuksia käyttää oikeuksiaan.<sup>313</sup>

Ihmisen suorittamaa valvontaa koitetaan tekoälysäädöksessä tehostaa lisäämällä valvontaa. Riittävän ihmisvalvonnan takaamiseksi biometriseen etätunnistukseen käytettävien tekoälyjärjestelmien tulokset tulee vahvistaa kahden luonnollisen ihmisen toimesta.<sup>314</sup> Kirjallisuudessa vaatimusta on kutsuttu neljän silmän periaatteeksi.<sup>315</sup> Säädöksessä ei oteta kantaa tilanteisiin, jossa vahvistavat luonnolliset henkilöt olisivat keskenään erimieltä järjestelmän tuottamasta tuloksesta. Kohdasta voidaan kuitenkin päätellä, että ihmisvalvonnan ei luoteta

---

<sup>310</sup> Enarsson et al. 2021 s. 123.

<sup>311</sup> Jones 2017 s. 14-15. Jones kutsuu tätä "oikeudeksi ihmisen osallisuuteen päätöksenteossa" (eng. "right to a human in the loop"). Lähtökohtana GDPR:ssä on, että ihmisellä on oikeus olla joutumatta sellaisen automaattisen päätöksenteon kohteeksi, joka perustuu täysin automaattiseen käsittelyyn, kuten profilointiin, ja jolla on häntä koskevia oikeusvaikutuksia. Pääsääntöön on kuitenkin poikkeuksia.

<sup>312</sup> Euroopan parlamentin & neuvoston asetus (EU) 2016/679 artikla 22.

<sup>313</sup> Tekoälysäädöksen johdanto-osan perustelukappale 171.

<sup>314</sup> Tekoälysäädöksen artikla 14(5).

<sup>315</sup> Ks. esim. Ruschemeier & Hondrich 2024 s. 22.

olevan täysin varma oikeussuojakeino, ja siksi vain yhden ihmisen suorittamana valvonta ei ole aina riittävä.

Oikeussuoja-aspektin lisäksi ihmisvalvonta toteuttaa oikeudellisesta näkökulmasta myös toista funktiota. Velvoite säilyttää vastuun valvovalla ihmisellä, jolloin ihmisellä voidaan sanoa olevan instrumentalistinen rooli vuorovaikutuksessa. Ihminen on väline vastuun toteuttamiselle kolmanteen nähden, vaikka järjestelmä olisi ihmiseen nähden aktiivisempi päätöksentekijä.<sup>316</sup> Tämän on huomautettu olevan hyvin ongelmallista vastuun syntymistä koskevien velvoiteoikeudellisten oppiemme kannalta. Vastuun syntymiseksi tulisi voida osoittaa moitittava toimi, jossa toimijalla on ollut mahdollisuus toimia toisin.<sup>317</sup> Näin ollen ihmisvalvonnan vastuuaspekti on yksi oikeuden digitalisaation haasteista.

Ihmisvalvonnan onnistumisiin tai epäonnistumisiin ja siitä seuraaviin vastuukysymyksiin liittyvä problematiikka on monimutkainen, laaja ja paljon käsitelty kokonaisuus. Käsittelyyn on sisältynyt sekä vastuun kohdentumiseen että vastuun kohdentumisen legitiimiyteen liittyviä kysymyksenasetteluita.<sup>318</sup> Vastuukysymykset ovat tämän tutkielman rajauksen ulkopuolella. Totean kuitenkin, että johdannossa esitetyllä tavalla yksi tekoälysäädöksen tavoitteista oli selventää vastuunkohdentumista tekoälyjärjestelmien arvoketjuissa. Siten tekoälysäädöksessä on paljon vaatimuksia, joiden avulla vedetään rajaa ihmisvalvojan ja kehittäjän vastuun välille.<sup>319</sup> Vaikka rajat vaikuttaisivat säädöksessä selvältä, nähtäväksi jää, millaiseksi oikeustila ihmisvalvonnan ja vastuunkohdentumisen kannalta muodostuu.

Toisaalta ihmisvalvonnan on oikeusyhteiskunnallisessa keskustelussa esitetty toteuttavan osittain yhteiskunnan fundamenttisimpia periaatteita kuten legitiimiyttä, jota ihmisen osallisuuden vähentyminen tai puuttuminen heikentää.<sup>320</sup> Toisin sanoen ihmisvalvonnan omaksumisen taustalla ei ole vain ihmisen instrumentalistinen rooli vastuun välittämisessä ja oikeussuojakeinona.<sup>321</sup> Lainsäädännön näkökulmasta ei

---

<sup>316</sup> Viljanen 2017 s. 1071-1073.

<sup>317</sup> Ibid.

<sup>318</sup> Constantino 2022 s. s.24 & 49; Malgieri 2019 s. 187; Veale & Zuiderveen Borgesius 2021 s. 103.

<sup>319</sup> Valvojaksi asetetun ihmisen tulee ainakin noudattaa järjestelmän tarjoajan käyttöohjeita ja tällä tulee olla riittävä koulutus. Lisäksi läpinäkyvyysvaatimukset ovat osa rajanvetoa. Tekoälysäädöksen johdanto-osan perustelukappale 72 ja artikla 13(2).

<sup>320</sup> Ks. esim. Laux 2023 s. 1; Koulu 2020 s. 734.

<sup>321</sup> Koulu 2020 s. 729.

kuitenkaan ole ihme, että ihmisvalvonta on omaksuttu. Ihminen on konkreettinen, oikeudellisesti tunnistettavissa ja tunnustettavissa oleva tuttu sääntelyn kohde.<sup>322</sup> Ihmisen ja järjestelmän vuorovaikutuksesta syntyy kuitenkin ainutlaatuinen vuorovaikutussuhde, joka on enemmän kuin tekijöidensä summa, eikä ole välttämättä säänneltävissä perinteisin menetelmin tai ihmisen kautta.<sup>323</sup> Ihmisen osallisuuden säilyttämiseksi olisi hyvä pohtia muita kuin ihmisvalvontaan perustuvia malleja, jotta ihmisen säilyttämisellä suojeltavat arvot eivät vaarannu.<sup>324</sup> Vähintäänkin on esitetty, että ihmisvalvonnan toteutus tulisi suunnitella ihmisvalvonnan todettujen heikkouksien kautta.<sup>325</sup> Tällä hetkellä tekoälyssäädöksessä valvovalle ihmiselle ei ole todettu muuta heikkoutta kuin taipumus luottaa liikaa automaattisten järjestelmien tuottamaan tulokseen (eng. “automation bias”).<sup>326</sup> Tekoälyssäädökseen omaksutun lähestymistavan riskinä on, että ihmisvalvonnan vaatimuksella muodostettu suojakaide ihmisten ja heidän intressien suojelemiseksi voi valvonnan vaikuttavuuden vähäisyyden vuoksi jäädä helminauhaksi.<sup>327</sup> Tämä ilmentää hyvin jakson alussa tunnistettuja teknologiasääntelyn haasteita, joista yksi oli sääntelyn tehottomuus.<sup>328</sup> Tehottoman sääntelyn legitiimiys voidaan asettaa kyseenalaiseksi, jolloin sääntely toimii tavoitettaan vastaan.

### 5.1.2 Ihmisen toimijuus luotettavan tekoälyn edellytyksenä – Järjestelmä välineenä

Kuten todettu, tekoälyssäädöksen politiikkadokumenteissa omaksuttiin lähtökohta, jonka mukaan tekoälyn tulee toimia työvälineenä ihmiselle (eng. “serve as a tool”).<sup>329</sup> Sama vaatimus tekoälyjärjestelmistä työvälineenä on omaksuttu tekoälyssäädöksessä, jossa isoin painotus onkin järjestelmän suorittamalla valmistelevilla ja hallinnollisilla tehtävillä.<sup>330</sup> Yleisesti tällaisista tehtävistä puhuttaessa käytetään termejä kuten “työväline” tai “edistävä” (eng. “tool tai ancillary”).<sup>331</sup> Esimerkiksi tekoälyjärjestelmät, joita käytetään työvälineinä poliittisten kampanjoiden hallinnollisessa ja logistisessa

---

<sup>322</sup> Crootof et al. 2023 s. 437.

<sup>323</sup> Ibid. s. 437-440.

<sup>324</sup> Koulu 2020 s. 734.

<sup>325</sup> Laux 2023 s. 8-11.

<sup>326</sup> Tekoälyssäädöksessä edellytetään, että valvova ihminen on tietoinen taipumuksesta ylikuottaa järjestelmän tuloksiin. Ks. Tekoälyssäädöksen 14(4(b)) artikla.

<sup>327</sup> von der Leyen (2023) käytti puheessaan suojakaide (eng. “guardrails”) vertausta puhuessaan tekoälyasetuksen tavoitteista.

<sup>328</sup> Brownsword 2008, s. 132.

<sup>329</sup> Tekoälyssäädöksen johdanto-osan perustelukappale 6.

<sup>330</sup> Ks. Tekoälyssäädöksen johdanto-osan perustelukappale 61.

<sup>331</sup> Esimerkiksi tekoälyssäädöksen johdanto-osan perustelukappaleet 27, 59, 61 & 62.



organisoimisessa, optimoinnissa tai strukturoimisessa eivät ole riskialttiita, sillä luonnolliset henkilöt eivät altistu välittömästi järjestelmän tuotokselle.<sup>332</sup> Lisäksi jo mainittu kieltä käyttää tekoälyjärjestelmiä luonnollisten henkilöiden rikoksentekoalttiuden arviointiin perustuen profilointiin tai persoonallisuuspiirteiden arviointiin ei sovellu järjestelmiin, joita käytetään ihmisarvioinnin tukena tilanteissa, joissa yksilön osallisuus rikolliseen toimintaan on jo objektiivisesti ja varmasti vahvistettavissa.<sup>333</sup> Vuorovaikutuskäsityksen vertailujakson mukaisesti, tämä vahvistaa vuorovaikutuksen dikotomista luonnetta.

Huolimatta tekoälyjärjestelmien roolista työvälteenä tekoälysäädöksessä on jo kuvattulla tavalla tunnistettu niiden mahdollisuudet huomattavaankin autonomiaan ja ympäristöään muovaaviin vaikutuksiin, jotka HCI-tutkimuksen mukaan muodostavat toimijan käsitteen.<sup>334</sup> Paikoitellen tekoälyjärjestelmien riskien kuvailussa käytetty kieli synnyttää mielikuvia merkittävänkin autonomian olemassaolosta ja siten uhasta ihmisen toimijuudelle. Näin on etenkin yleiskäyttöisten tekoälymallien (eng. “General Purpose AI, GPAI”) osalta, joihin voi liittyä niin sanottu järjestelmätason riski (eng. “systemic risk”). Yleiskäyttöisten tekoälymallien käsite haluttiin erottaa tekoälyjärjestelmien käsitteestä, sillä yleiskäyttöisille tekoälymalleille ja järjestelmätason riskin muodostaville tekoälymalleille asetettiin omia määräyksiä, jotka tulevat sovellettavaksi riippumatta, ovatko mallit tekoälyjärjestelmien komponentteja vai eivät.<sup>335</sup> Yleiskäyttöisten mallien määritelmä perustuu mallien tyypillisimmille ominaisuuksille, joita ovat mallien yleisyys ja kyky pätevästi suorittaa laaja-alaisia ja moninaisia tehtäviä (eng. “wide range of distinctive tasks”).<sup>336</sup> Mallit ovat tyypillisesti koulutettu suurella määrällä dataa itseohjatusti.<sup>337</sup> Järjestelmätason riskillä tarkoitetaan tosiasiallisia tai kohtuudella ennakoitavissa olevia kielteisiä vaikutuksia esimerkiksi terveydelle ja turvallisuudelle, demokraattisille prosesseille, kriittisille aloille tai laittoman, väärän

---

332 Tekoälysäädöksen johdanto-osan perustelukappale 62.

333 Tekoälysäädös artikla 5(1(d)).

334 Esim. Sterling & Taveter 2009.

335 Tekoälysäädöksen johdanto-osan perustelukappale 97. Järjestelmän ja mallin erottaa käyttöliittymä, joka mahdollistaa mallin käyttämisen ja vuorovaikutuksen syntymisen järjestelmän ja käyttäjän välille.

336 Ibid. Suuret generatiiviset tekoälymallit ovat esimerkki yleiskäyttöisestä tekoälymallista. Tehtäviä, joita nämä mallit suorittavat mahdollistavat joustavan sisällön luomisen esimerkiksi kuvan, tekstin, äänen tai videon muodossa, joka voidaan mukauttaa/sovittaa laaja-alaisesti monenlaisten tehtävien osaksi. Tekoälysäädöksen johdanto-osan perustelukappale 99.

337 Tekoälysäädöksen johdanto-osan perustelukappale 97 & artikla 3(63).

ja syrjivän sisällön leviämistä.<sup>338</sup> Järjestelmätason riskin katsotaan muodostuvan järjestelmän ominaisuuksien kehittyessä. Olennaista on ominaisuuksien kokonaisvaltainen arviointi, jossa huomioidaan esimerkiksi järjestelmän luotettavuus, turvallisuus, autonomian aste, sen pääsy eri välineisiin, mahdollisuus poistaa turvajärjestelmiä ja mahdollisuudet tahalliseen tai tahattomaan väärinkäyttöön.<sup>339</sup>

Yleiskäyttöisten tekoälymallien ja järjestelmätason riskiä käsittelevistä kohdista huomataan, että autonomia ja ihmisen toimijuudesta irtaantuminen muodostaa uhan. Tällöin syntyy vaikutelma, jonka mukaan uhat ovat tekoälyjärjestelmien ominaispiirteitä, jotka eivät muodostu ihmisten tavasta käyttää tai suunnitella järjestelmiä, vaan ovat sisäänrakennettuja tekoälyn ominaisuuksia. Tämä ajattelutapa toistaa STS-tutkimuksessa käytyä keskustelua siitä, onko ihmisten luomilla artefakteilla oma politiikkansa.<sup>340</sup> Ajattelumalli johtaa helposti teknologiadeterminismiin, jonka mukaan uusi teknologia tuo aina mukanaan automaattisia vaikutuksia, joihin ei voida yhteiskunnallisesti vaikuttaa. Tekoälysäädöksen politiikkadokumenteissa huomattiin sama ilmiö riskikuvailun osalta. Tässä kontekstissa onkin huomautettu, että uhkien maalaaminen tekoälyjärjestelmien ominaispiirteiksi ei ole hedelmällistä pohdittaessa tarkoituksenmukaisia sääntelyratkaisuja.<sup>341</sup> Siten epätarkka keskustelu uhkien muodostajista vaikeuttaa tehokkaan sääntelyn luomista. Se vaikuttaa kielteisesti myös mahdollisuuteen luoda yhteistyöhön perustuvaa käsitystä vuorovaikutusta. HCI-tutkimuksen tarkoittamaa “me-toimijuutta” tuskin syntyy, jos vuorovaikutuksen toinen osapuoli koetaan uhkana.

Yleismallien ja järjestelmätason riskiä käsittelevien kohtien lisäksi tekoälyjärjestelmien riskeistä yleisesti puhuttaessa toistuu sama havainto.<sup>342</sup> Lisäksi jo todetulla tavalla sekä tekoälyn määritelmä että osa tekoälyjärjestelmiä koskevista vaatimuksista tähdätään juuri niiden ominaisuuksien mitikointiin, joita HCI-tutkimuksessa tunnistettiin korkean autonomian järjestelmille. Säädöksessä ei kuitenkaan puhuta autonomian ja ihmisestä riippumattomuudesta yhteydessä

---

338 Tekoälysäädöksen johdanto-osan perustelukappale 110 & artikla 3(65).

339 Tekoälysäädöksen johdanto-osan perustelukappale 111.

340 Winner 1980 s. 121-136. Poliitiikalla kirjoituksessa tarkoitetaan vallan ja auktoriteetin järjestämistä ihmisyhteisössä ja näissä tapahtuvaa toimintaa.

341 Ks. Koulu 2020 s. 35-39.

342 Tekoälysäädöksen johdanto-osan perustelukappaleet 47, 53 & 64.

järjestelmän toimijuudesta. Ylläesitetyllä tavalla säädös vaikuttaa ristiriitaiselta, koska etenkin riskikuvausten yhteydessä esitetty tekoälyjärjestelmien toimijuuden aste ei vastaa niille käsitteellisesti annettua oikeudellista asemaa työvälineenä.

Nähdäkseni tekoälyjärjestelmien rooli työvälineenä oikeudellistuvassa vuorovaikutuskäsityksessä voidaan selittää toimijoihin oikeustieteessä liitettävällä käsitejärjestelmällä. Siinä missä HCI-tutkimus lähestyy toimijuutta suunnittelun näkökulmasta, oikeustiede kytkee sen lopulta oikeussubjektin käsitteeseen.<sup>343</sup>

Kansainvälisessä varallisuus-oikeudessa oikeussubjektisuuden käsitettä onkin lähestytty kolmijaon kautta, joka käsittää välineet (eng. “tools”), edustajat (eng. “agents”) ja henkilöt (eng. “persons”).<sup>344</sup> Selvää on, että EU-oikeus ei pidä tekoälyä tai tekoälyjärjestelmiä oikeussubjekteina.<sup>345</sup> Keskustelu mahdollisuudesta ulottaa oikeussubjektius tekoälylle on alkanut teoreettisella tasolla jo vuonna 1981. Vaikka Lehman-Wilzing avasi keskustelun viittauksella Frankensteinin hirviöön, kirjoitus pitää sisällään samoja oikeudellisia analogioita, joita tekoälyjärjestelmien paikantamiseen oikeusjärjestelmässä nykyäänkin liitetään. Hän nostaa esiin muun muassa tuotevastuun, edustuksen (eng. “agent”) ja viimeisimpänä oikeussubjektisuuden (eng. “person”).<sup>346</sup> Toisen klassikkoteoksen aiheesta on kirjoittanut Lawrence B. Solum vuonna 1992.<sup>347</sup> Hän ei pitänyt kysymyksenasettelua oikeussubjektuudesta enää puhtaana teorisointina ja leikittelynä ajatuksella, vaan penäsi vakavasti otettavan keskustelun aloittamista korostamalla oikeuden yhteiskunnallista merkitystä.<sup>348</sup> Keskustelu oikeussubjektisuuden myöntämisestä on edennyt pisteeseen, jossa tunnustetaan, ettei monessakaan oikeusjärjestelmässä olisi esteitä luoda uutta oikeussubjektisuuden kategoriaa tai uudistaa käsitteitä siten, että jonkinasteinen oikeussubjektius olisi myönnettävissä.<sup>349</sup> Kirjoitukset eivät

---

343 Oikeudelliset käsitteet oikeussubjektuudesta puhuttaessa vaihtelevat oikeuskulttuurien ja oikeusjärjestelmien välillä. Etenkin suomalaisessa oikeustieteessä käsite oikeussubjekti ymmärretään oikeuskelpoisuuden kautta. Ks. esim. Kurki 2018 s. 822. Oikeussubjektia käytetään taasen helposti henkilö-sanalla synonyymienä, se kattaa luonnolliset- ja oikeushenkilöt. Kurki 2018 s. 470.

344 Kurki 2018 s. 820.

345 Vrt. Euroopan parlamentti 2017 (2015/2103(INL)).

346 Lehman-Wilzing 1981 s. 422-457. Alun perin termi “robot” on johdettu tsekin kielen sanasta ”robota”, joka kuvaa työvoimaa ja pakkotyötä. s. 449.

347 Solum 1992 s. 1234-1240.

348 Solum 1992 s. 1232-1233. Solum korosti käsitteiden merkitystä keskustelun kannalta (eng. “The law embodies core insights about the way the world works and how we evaluate it.”) s. 1232. Hänen kirjoituksensa ensimmäinen osa keskittyi esittelemään kognitiotieteiden tuloksia, jotka hän kytki oikeussubjektin käsitteeseen.

349 Ks. Kurki 2018 s. 820-839. Havainto sopii hyvin yhteen modernin oikeuden instrumentalistisen luonteen kanssa.

kuitenkaan ota kantaa siihen, tulisiko tällaista uudistusta edistää.<sup>350</sup> Esimerkiksi Euroopan parlamentin komissiolle suuntamaan päätöslauselmaan saatu kritiikki paljastaa, että keskustelussa liikutaan olemassa olevien kategorisointien välillä, jolloin aiheeseen otetaan kantaa “joko tai”-tyyppisesti.<sup>351</sup> Niissä keskitytään siten vertailemaan, pitäisikö tekoälystä tulla ihmisen kaltainen luonnollinen henkilö, oikeushenkilö vai kumpikaan näistä.<sup>352</sup>

Oikeussubjektius on luonnollisesti hyvin keskeinen ja merkityksellinen oikeudellinen konstruktio. Keskustelu oikeussubjektuudesta on hyvin teoreettista ja sitä ei ole mahdollista käydä syvällisesti tässä tutkielmassa. Haluan kuitenkin vertailla hieman oikeudellista toimijan käsitettä HCI-tutkimuksen vastaavaan ja arvioida, olisiko tavoitteiden kautta ymmärrettävä toimijan käsite mahdollinen myös oikeudellisessa kontekstissa. Tämä voisi puolestaan johtaa siihen, että järjestelmät voitaisiin nähdä edustajina (eng. “agents”) välineiden sijaan kuten HCI-orientaatioissa. Joka voisi vuorostaan johtaa siihen, että oikeudellisen vuorovaikutuskäsityksen dikotomisuus vähenisi ja vuorovaikutuksen yhteistyöulottuvuus korostuisi.

Oikeussubjektuuden saavuttamiseksi toimijalta katsotaan vaadittavan lähestymistavasta riippuen joko kykyä rationaaliseen ja autonomiseen toimintaan (tahtoteoria) tai toimijalle myönnettävää perimmäistä arvoa (intressiteoria).<sup>353</sup> Järjestelmät eivät ole perimmäisesti arvokkaita mutta jo todetulla tavalla tekoälyjärjestelmät ovat jossain määrin autonomisia ja kykenevät vuorovaikuttamaan

---

350 Ibid. Nähdäkseni keskustelu tulisikin kääntää nimenomaan tähän ulottuvuuteen. Keskeinen kysymys lienee, millaisia hyötyjä ja haittoja lähestymistavat pitää sisällään. En tarkoita tällä täysin praktisen näkökulman omaksumista, jonka mukaan oikeudellisen subjektin (“person”) käsite on pelkkä konstruktio, jolla ei ole kiinteää sisältöä ja jolle lainsäädäntö mahdollistaa minkä tahansa käytännössä toteutettavan mahdollisuuden. Vaan pidän tärkeänä käsitteen kytkemistä jollain tasolla toimijuuteen ja vastuuntoteuttamiseen. Ks. Thomas 2011 s. 631-646. Rationaalisen lähestymistavan mukaan olennaista on kapasiteetti järkeilyyn, perusteluun (reasoning) ja valintaan. Tästä muodostuu olennaisin oikeudellisen subjektuuden sisältö.

351 Ks. Euroopan parlamentti 2017 (2015/2103(INL)) ja saatu kritiikki: Open letter to the European Commission Artificial Intelligence and Robotics.

352 Ehdotuksen kritisoitiin ensinnäkin yliarvioivan älykkäimpienkään järjestelmien edistyksellisyys. Toiseksi mahdollisuudet juontaa sähköinen henkilöllisyys olemassa olevista oikeussubjektuuden lähestymistavoista poissuljettiin jotakuinkin kapein perustein: luonnollisen henkilöyden kritisoitiin olevan vastoin EU:n perusoikeuskirjaa ja Euroopan ihmisoikeussopimusta (rationaalinen painotus); oikeushenkilöllisyyden katsottiin vaativan luonnollisen henkilön osallisuutta edustajana ja johtajana; samoin anglosaksisen trust-instituution katsottiin vaativan viimekädessä vastuunkohdistumista ihmiseen. Ks. Open letter to the European Commission Artificial Intelligence and Robotics.

353 Lähestymistavat käsitteeseen riippuvat selityksestä oikeuksien funktiolle. Subjektivisten oikeuksien funktioita voidaan selittää tahto- ja intressiteorialla. Molemmat teoriat keskittyvät oikeuksien selittämiseen niiden haltijan näkökulmasta. Aalto-Heinilä 2009 s. 138-139.

ympäristönsä kanssa.<sup>354</sup> Oikeuden kontekstissa ihmisen perimmäinen arvokkuus määritellään autonomian kautta.<sup>355</sup> Autonomia ymmärretään yksilön kyvyksi määrätä asioistaan itse. Autonomiia käytetään joskus itsemääräämisoikeuden synonyymina, vaikka itsemääräämisoikeus viittaa vain oikeuteen määrätä asioistaan.<sup>356</sup> Autonomia ja itsemääräämisoikeus voidaan oikeustieteessä ulottaa yksilön lisäksi ryhmiin.<sup>357</sup> HCI-tutkimuksessa autonomian todettiin olevan toimijuuden osatekijä. Myös HCI-orientaatiossa autonomialla viitattiin kykyyn luoda oma agenda vailla toisen toimijan valtaa. Autonomialla on siten osin yhteinen merkitys tieteiden välillä. Kuitenkin HCI-tutkimuksessa toimijan käsitteen kannalta autonomiia tärkeämmäksi tekijäksi nostettiin kyky vuorovaikuttaa ympäristön kanssa ja vaan riittävää on, että agentti toimii sille asetetun tavoitteen toteuttamiseksi suhteessa ympäristöön tavalla, joka ilmentää kykyä proaktiivisuuteen ja reaktiivisuuteen.<sup>358</sup> Siten HCI-tutkimuksen mukaan järjestelmän toimijuuden kannalta autonomisuutta ei tarvitse nähdä absoluuttisena, vaan se syntyy tavoiteorientaation kautta.

Tavoiteorientoitunutta lähestymistapaa voidaan käyttää oikeustieteessäkin. Visa Kurki esittelee tekoälyn oikeussubjektiutta käsittelevässä artikkelissaan funktionaalisen toimijuuden teorian, jossa hyödynnetään intentionaalisen suhtautumisen omaksumista. Funktionaalisen toimijuuden teoria on yleinen lähestymistapa ryhmien tutkimisessa. Teorian mukaan organisoitunut ihmisjoukko voidaan yksilön tapaan ymmärtää toimijaksi, vaikkei joukolle ole osoitettavissa subjektiivista mieltä. Toimijuuden arvioinnissa olennaista on, voidaanko joukolla katsoa olevan uskomuksia, tavoitteita ja aikomuksia. Intentionaalinen

---

354 HCI-tutkimuksessa tunnistettiin, että toimijan käsitellä tarkoitetaan oliota, joka suorittaa tiettyä toimintaa ympäristössä, josta se on tietoinen ja voi reagoida muutoksiin.

355 Autonomian toteuttamiseen liittyy olennaisesti vapaan tahdon problematiikka, joka on filosofisena kysymyksenä paljon kiistelty aihe. Ks. Tiensuu et al. 2023 s. 11. Vapaan tahdon metafysiseseen ulottuvuuteen on mahdoton ja jokseenkin turha ottaa tässä kontekstissa kantaa. Aihe on relevantti oikeudellisesta ja poliittisesta kontekstista. Nähdäkseni oikeudellinen konteksti ei edellytä vastaamista tähän, koska oikeusnormisto vapauttaa osittain vapaan tahdon edellytyksestä. Lainsäädäntö tarjoaa tietyn moraalinnormiston valmiina, toimija noudattaa tätä normistoa monesti omaksuen sen. Toisin sanoen vastuullisuuteen riittää, että toimija kykenee tunnistamaan normien olemassaolon ja mukauttamaan vuorovaikutustaan ympäristön kanssa siten, että normeja ei rikota. Autonomiia kuvaa toimijan osuus, joka jää normien tunnistamisen ja toiminnanohjauksen väliin.

356 Tiensuu et al. 2023 s. 9-10.

357 Ibid s. 10-11. Yhteiskunnallisesta näkökulmasta autonomia edellyttää mahdollisuutta osallistua itseä koskevaan päätöksentekoon. s. 11.

358 Autonomiia tärkeämpää on toimijan paikallisuus ympäristössään. Tähän kuului esimerkiksi kyky havainnoida, vastaanottaa ja lähettää viestejä sekä vaikuttaa ympäristöön. Älykkäällä toimijalla näihin kykyihin kuuluu lisäksi reaktiivisuus ja proaktiivisuus. Sterling & Taveter 2009 s. 38.

suhtautuminen avulla voimme mieltää toimijaksi olion, jonka toimintaa voidaan selittää parhaiten sen tavoitteen kautta riippumatta, onko sillä mielentiloja tai ajatuksia. Tavoitteiden ei tarvitse palautua yksittäiseen toimijaan, vaan ne voivat olla johdettavissa esimerkiksi organisaatorakenteesta tai päätöksentekojärjestelmistä.<sup>359</sup>

Funktionaalisen toimijuuden teoria on nähdäkseni yhteensopiva kansainvälisen oikeuden kanssa, jossa selkeästi tunnustetaan ihmisryhmien mahdollisuus autonomiaan.<sup>360</sup> Toki erona esimerkiksi ihmisen ja järjestelmän muodostamalle yksikölle on se, että nämä ihmisryhmät koostuvat perimmäisesti arvokkaista yksilöistä. Aineiston analyysijaksossa tuotiin esiin, että järjestelmän ominaisuudet ovat järjestelmän tavoitteita. Tästä näkökulmasta järjestelmiin voisi suhtautua myös oikeudellisesti toimijoina, vaikka niillä ei ole oikeussubjektiuteen liitettävää perimmäistä arvoa tai ne eivät ole autonomisia siinä mielessä, että tavoite olisi järjestelmän itse määrittämä. Myös oikeuden näkökulmasta tekoälyjärjestelmät voitaisiin siis ymmärtää pikemminkin edustajina, koska niillä vaikuttaa tekoälysäädöksen mukaan olevan (implisiittinen) toimijuus, joka tekee niistä enemmän kuin työvälineitä muttei kuitenkaan oikeussubjekteja. Jos oikeudellinen vuorovaikutuskäsitys lähestyisi järjestelmiä edustajina, joille ihmiset määrittelevät tavoitteet, vuorovaikutuskäsitys ei olisi niin dikotominen. Vaikka järjestelmien asema edustajana toisi uusia oikeudellisia kysymyksiä ratkaistavaksi se ei välttämättä muuttaisi vastuuasemia.<sup>361</sup> Vastuu säilyisi ihmisellä, jonka intressejä järjestelmä edustaa.<sup>362</sup> Tokikaan käsitys järjestelmistä edustajina ei ole yhteensopiva säädöksessä käytetyn riskikuvailun kanssa, sillä edustajien oletetaan toimivan päämiehensä edun mukaisesti. Riskien osalta ihmistoimijan osuus on säädöksessä osittain häivytetty, ja juuri se legitimoii osittain sääntelyn tarvetta. Muutos olisi kuitenkin käsitteellinen ja se vähentäisi tekoälysäädöksen sisäistä ristiriitaisuutta.

Tavoitteisiin pyrkiessä myös agentin tosiasialliset vaikutukset ympäristöön määrittävät HCI-tutkimuksen mukaan olennaisesti toimijuuden sisältöä. HCI-tutkimuksessa perinteinen ihmisen ja teknologian instrumentaalinen suhde haluttiin

---

<sup>359</sup> Kurki 2018 s. 821.

<sup>360</sup> Vrt. (ILO) Yeissopimus alkuperäiskansojen ja heimokansojen oikeuksista riippumattomissa maissa (Nro 169) ja YK:n julistuksesta alkuperäiskansojen oikeuksista.

<sup>361</sup> Hildebrandt 2020 s. 248.

<sup>362</sup> vrt.kritiikki EP:n päätöslauselmaan (2015/2103(INL)) saatu kritiikki: Open letter to the European Commission Artificial Intelligence and Robotics.

korvata “me (ihminen ja teknologia) identiteetillä”, koska teknologian tosiasiallinen vaikutus ympäristöä muovaavana tekijänä havaittiin. Instrumentaalisella suhteella viitataan dikotomiaan, jossa ihmisen toiminnot nähdään riippumattomana teknologiasta ja teknologia työvälineenä eli instrumenttina. Uusi filosofia ammennettiin valistuksen dialektiikan lähtökohdista, jonka mukaan ihminen toiminnallaan muokkaa ympäristöään, joka yhä muokkaa ihmistä. Ihmisen ja teknologian vuorovaikutussuhteeseen siirrettynä tällä tarkoitettiin yhteistyön korostamista, jossa ihmisen ja teknologian suhde on kehittymässä pelkkää vuorovaikutusta pidemmälle suhteeksi, jossa molemmat toimijat osallistuvat ihmisyyden muokkaamiseen.<sup>363</sup>

Samoin tekoälyseudöksessä tunnistetaan selkeästi järjestelmien tosiasiallinen vaikutus ympäristöön, mutta jo todetulla tavalla instrumentaalisesta suhteesta ei luovuta. Sääöksessä on monissa konteksteissa tunnistettu järjestelmien vaikutukset ihmisen toimijuutta ja toimintaa muokkaavana tekijänä. Kiellettyjen subliminaalisten tapojen lisäksi muutos voi olla hienovaraisempaa ja hyväksyttävämpää. Esimerkiksi tekoälyjärjestelmien käytön katsotaan olevan tärkeää koulutuksessa, jotta ihmisille voidaan varmistaa riittävät digi- ja medianlukutaidot.<sup>364</sup> Nämä ovat välttämättömiä nyky-yhteiskuntaan osallistumisen kannalta ja siten oppiminen vuorovaikutukseen on välttämätöntä aktiiviseksi kansalaiseksi kasvamisessa. Vastaavasti luovalla alalla tapahtunutta sosiaalista muutosta kuvataan punnitsemalla toisaalta generatiivisten mallien mahdollistamilla innovaatioilla, jotka saman aikaisesti haastavat kirjailijoiden, muusikoiden ja muiden sisällön luoja tavan luoda, jaella, käyttää ja kuluttaa teoksiaan, vaikuttaen esimerkiksi tekijänoikeuksiin.<sup>365</sup> Edeltävät esimerkit osoittavat, kuinka moninaisia vaikutuksia tekoälyjärjestelmien ja ihmisen vuorovaikutuksella nähdään olevan. Juuri tekoälyteknologian vuorovaikutteisuuden vuoksi instrumentalistisesta lähestymistavasta voisi etäännyä. Samalla tunnustettaisiin järjestelmien tosiasiallinen vaikutus ympäristössä. Uudet tilanteet edellyttävät kiistämättä oikeudellisten normien uudelleen muotoilua, jotta ihmistoimijaa- ja toimintaa voidaan suojata. Toisaalta vuorovaikutus on välttämättömyys, joka myös aktivoi ja mahdollistaa oman toimijuuden toteuttamisen.

---

363 Cho & Yoon 2013 s. 32-35.

364 Tekoälyseudöksen johdanto-osan perustelukappale 56. Sen sijaan koulutuspolkuun esimerkiksi arvioinnein vaikuttavat järjestelmät muodostavat korkean riskin, koska voivat vaikuttaa ihmisen toimijuuteen rajoittavasti.

365 Tekoälyseudöksen johdanto-osan perustelukappale 105.

Kohdat kuvaavat hyvin myös teknologian, yhteiskunnan ja oikeuden dynaamista vuorovaikutussuhdetta.

### 5.1.3 Ihmisen päätöksenteko luotettavan tekoälyn edellytyksenä

Tekoälysäädöksessä päätöksentekoon liittyvä roolitus ihmisen ja järjestelmän välillä on oikeudellisessa ja merkittäviä oikeudellisia vaikutuksia saavassa päätöksenteossa jyrkkä. Vuorovaikutuksessa molemmille on selkeästi osoitettavissa omat tehtävät, jossa järjestelmän tehtävät voivat koostua lähinnä valmistelevista ja päätöksen sisältöön vaikuttamattomista tehtävistä. Ihmisen rooli on taas olla tietoinen ja suhteellisen suvereeni päätöksentekijä. Tekoälysäädöksessä luetellaan esimerkkien avulla, kuinka järjestelmiä voidaan käyttää oikeudellisessa päätöksenteossa. Esimerkiksi oikeuden ja demokraattisten prosessien yhteydessä hallinnollisia tehtäviä edistävät toimet kuten oikeudellisten päätösten, dokumenttien, datan ja kommunikaation anonymisointi ja pseudonymisointi sekä henkilöstön kommunikointi voidaan suorittaa tekoälyjärjestelmän toimesta, ilman korkeaa riskiä.<sup>366</sup> Lisäksi säädöksessä on kriteeristö, jonka avulla voidaan arvioida, onko tekoälyjärjestelmä vaikuttanut olennaisesti päätöksen lopputulokseen. Mikäli yksi tai useampi kriteereistä täyttyy, olennaista vaikutusta tuskin on. Kriteerien mukaan tekoälyjärjestelmän käyttötarkoituksena pitäisi olla kapean proseduraalisen tehtävän täyttö. Tällaisia ovat esimerkiksi strukturoimattoman datan strukturointi, saapuvien dokumenttien järjestely ja luokittelu sekä kaksoiskappaleiden tunnistaminen tietomassoista. Vaikka järjestelmää käytettäisiin säädöksen liitteessä III määritellyissä korkean riskin toiminnoissa, tällaiset tehtävät eivät muodosta uhkaa niiden kapea-alaisuuden ja rajoittuneen luonteen vuoksi.<sup>367</sup> Toiseksi kriteeri olennaisen vaikutuksen puuttumisesta voi täytyä, mikäli tekoälyjärjestelmän tehtävä on parannella aiempaa ihmisen tekemää työtä. Tällaisella parantelulla viitataan esimerkiksi tekstin hiomiseen vastaamaan tiettyä akateemista tai ammattimaista sävyä tai yhdenmukaistamaan tekstiä kielellisesti.<sup>368</sup> Kolmanneksi kriteeri olennaisen vaikutuksen puuttumisesta voi täytyä, mikäli tekoälyjärjestelmän käyttötarkoitus on havaita kaavamaisuuksia aiemmasta ihmisen suorittamasta päätöksenteosta tai poikkeamia havaitusta kaavasta. Tällaisessa toiminnossa ei nähdä riskiä, sillä

---

<sup>366</sup> Tekoälysäädöksen johdanto-osan perustelukappale 61.

<sup>367</sup> Tekoälysäädöksen johdanto-osan perustelukappale 53 & artikla 6(3).

<sup>368</sup> Tekoälysäädöksen johdanto-osan perustelukappale 53.



tekoälyjärjestelmä seuraa ihmisen aiemmin suorittamaa päätöksentekoa, eikä sitä pyritä korvaamaan tai siihen ei pyritä vaikuttamaan. Esimerkkinä tällaisista tehtävistä on esimerkiksi opettajan arvostelulinjan yhdenmukaisuuden varmistamiseen käytettävät tekoälyjärjestelmät.<sup>369</sup> Neljäs ja epämääräisin kriteeri on, että tekoälyjärjestelmä tekee vain sellaisia valmistelevia tehtäviä arvioinnin suorittamiseksi liitteen III mukaisissa käyttötarkoituksissa, joiden outputin vaikutusmahdollisuudet sitä seuraavaan arviointiin ovat vähäiset. Tällaisia sovellutuksia voi olla esimerkiksi tietojenhallintatyökalut, indeksointityökalut, hakutoiminnot, tekstin tai puheen prosessointi, datan linkittäminen toisiin datalähteisiin ja käänösavustimet.<sup>370</sup> Kuten voimme esimerkeistä huomata, oikeudellinen päätöksenteko pysyy tiukasti ihmisen suorittamana.

Ottaen huomioon järjestelmien muissa yhteyksissä tunnistettu kyvykkyys sekä HCI-tutkimuksen esittämä tavoiteltu vuorovaikutusmalli on olennaista kysyä, mikä tekee oikeudellisesta päätöksenteosta järjestelmälle ulottumattoman. Toisin sanoen, mikä oikeudellisen päätöksenteon piirteistä tekee siitä erityisen ja vain ihmiselle varatun.<sup>371</sup> Koska vastuukysymysten arviointi on tutkielman rajauksen ulkopuolella, en pohdi, mikä vaikutus esimerkiksi virkavastuulla asiaan.<sup>372</sup> Se on yksi jo mainituista oikeuden digitalisaation haasteista.

Tekoälyssäädöksessä todetaan, että tekoälyjärjestelmät voivat tukea tuomareiden päätöksentekovaltaa ja heidän oikeudellista itsenäisyytensä, mutta järjestelmät eivät saa korvata sitä.<sup>373</sup> Sen sijaan toiminta, joka ei vaikuta päätöksen asiasisältöön voi olla kumman tahansa toimijan – ihmisen tai järjestelmän tehtävissä.<sup>374</sup> Vaikutuksia päätöksen asiasisältöön arvioidaan yllä esitetyn kriteeristön avulla.

Oikeustieteellisessä kirjallisuudessa tekoälyjärjestelmien ja ihmisen yhteistoiminnallisesta päätöksenteosta käytetään hybridipäätöksenteon käsitettä.<sup>375</sup>

Oikeudellisen päätöksenteon mallintamisen haasteet on nostettu esiin oikeusinformatiikassa 1950-luvulta alkaen. Loevingerin mukaan haaste on

---

<sup>369</sup> Ibid.

<sup>370</sup> Ibid.

<sup>371</sup> Koulu 2018 s. 842 & 851. Tällä pyritään oikeuden itsetietoisuuden lisäämiseen ja se voidaan nähdä teknologian tuomien kysymyksenasetteluiden positiivisena ulottuvuutena.

<sup>372</sup> Aiheesta esim. Koulu 2020 720-735.

<sup>373</sup> Tekoälyssäädöksen johdanto-osan perustelukappale 61.

<sup>374</sup> Tekoälyssäädöksen johdanto-osan perustelukappale 53.

<sup>375</sup> Esim. Enarsson et al. 2022 s. 124.

oikeudellisen päätöksenteon kääntäminen koneluettavaan muotoon, joka johtuu erityisesti oikeuden luonteesta kielellisenä ilmiönä. Hänen mukaansa: “Oikeudelliset termit ovat lähes kaikki epämääräisiä sanamuotoja, joilla on vain rituaalinen merkitys.”<sup>376</sup> Toteamus pitää nähdäkseni sisällään kritiikkiä oikeudellista päätöksentekoa kohtaan. Ainakin siltä osin, kun päätöksenteon esitetään olevan loogista. Loevingerin mukaan oikeudelliset termit ovat tietyllä tavalla vain representaatio oikeudellisesta päätöksestä. Kun päätös on tehty, valitaan sopivin oikeudellinen käsite kuvaamaan toimintaa. Valinta oikeudellisen termin käytölle ei ole looginen, se on korkeintaan intuitiivinen.<sup>377</sup>

Vastaavasti varhainen kotimainen oikeuskirjallisuus asetti tietoteknologioiden käytön rajoitteeksi oikeudellisen ratkaisutoiminnan kompleksisuuden. Makkonen esitti 80-luvulla käännettyssä monografiassaan, että rajoitteena on ratkaisutoiminnan loogisen esittämisen mahdottomuus. Samalla hän korosti, että teknologioilla voidaan helpottaa ratkaisutoimintaa muutoin.<sup>378</sup> Loevingerin tapaan Makkonen päätyi korostamaan - joskin positiivisempaan sävyyn - ratkaisutoiminnan vaikeasti jäsennettäviä inhimillisiä ulottuvuuksia, jotka liittyvät ihmisen tiedonkäsittelyn tapoihin. Jorma Kuopus jatkoi mallintamisen vaikeutta koskevaa keskustelua esittämällä, että päätöksenteon määrälliset suureet voidaan toteuttaa automaatiolla mutta kvalitatiiviset arvot eivät ole käännettävissä binäärilogiikkaan.<sup>379</sup> Tähän liittyi huoli arvojen väkinäisestä kääntämisestä.<sup>380</sup>

Yllä esitetyllä tavalla oikeudelliseen päätöksentekoon kuuluu olennaisesti tietyn asteinen epävarmuus. Yleisesti tunnustetaan, että oikeudellinen päättely ei

---

376 Loevinger 1949 s. 472.

377 Ibid. Loevingerillä on selkeä empiriapainotus: “If mankind is capable of learning anything, it should have learned by this time that it can solve no problems by introspection, but only by investigation. Knowledge cannot be acquired by speculation, but only by observation.” s. 472.

378 Makkonen 1981 s. 253. Makkosen esittämällä tavalla teknologiaa on otettu käyttöön merkittävästi ratkaisutoiminnassa muutoin. Esimerkiksi yleisissä tuomioistuimissa toteutettiin juuri mittava asian- ja dokumentinhallinnan kehittämishanke (AIPA), jonka tarkoituksena oli kehittää sähköisiä menetelmiä. Ks. Oikeusministeriö 2015.

379 Älykkyyden englanninkielinen käännös intelligence on johdettu latinan kielen verbistä ”intelligere”, jonka sananmukainen käännös on ”valita vaihtoehtojen välillä” (inter – between, among, legere – to choose). Tämä määritelmä on yhtäläinen psykologiassa käytettävän mitattavan (quantifiable) älykkyyden käsitteen kanssa. Tällainen älykkyyden määritelmä on osunut saumattomasti yhteen 1960-luvulta lähtöisin olevaan käsitykseen tekoälystä, joka perustuu pitkälti matematiikkaan ja laskentatehoon. Ihmisälyn laadullista ulottuvuutta on vaikea mitata ja toisintaa. Ks. Žáčková 2011 s. 135.

380 Jorma Kuopus 1988, s. 505. Yhtäläisyyksiä nykyiseen inhimillisiä tekijöitä korostavaan keskusteluun esim. Hildebrandt 2019 s. 1.

välttämättä noudata logiikkaa.<sup>381</sup> Lisäksi oikeudelliseen päätöksentekoon liitetään olennaisesti inhimillisintuutiivinen ulottuvuus.<sup>382</sup> Nähdäkseni siten selvää on, että päätöksentekoprosessin ei tarvitsekaan olla looginen, oikeudellisten päätösten perusteiden tulee. Samalla hyväksytään, että ihmisenkin päätöksentekoprosessit ovat läpinäkyvämpiä. Kuitenkin ne tulevat hyväksyttäväksi oikeudellisesti uskottavilla perusteilla, jotka kirjoitetaan näkyväksi päätökseksi.

Automaation aikakaudella oikeudellisen päätöksenteon säilyttäminen ihmisellä perusteltiin siis etenkin oikeudellisen päättelyn loogisen esittämisen mahdottomuutena. Sama perustelu ei ole kelpo datavetoisten tekoälyjärjestelmien kohdalla, sillä esimerkiksi koneoppimista hyödyntämällä järjestelmät voitaisiin kouluttaa valtavalla määrällä oikeudellista aineistoa, joka mahdollistaa oikeudellisen päättelyn jäljittelyn.<sup>383</sup> Järjestelmät siis oppivat säännönmukaisuudet aineistosta. Tekoälynsäädöksen mukaan tekoälyjärjestelmät sääntöpohjaisista järjestelmistä erottaa juuri niiden kyky tehdä päätelmiä.<sup>384</sup> Teknologian kehitysaste näkyikin korostuneesti aiheita käsittelevässä oikeustieteellisessä keskustelussa.<sup>385</sup> Vaikka yllä olevat kuvaukset haasteista on esitetty verraten sääntöpohjaiseen päätöksentekoon, niitä voidaan hyödyntää arvioidessa, millaisena oikeudellisena päätöksentekijänä ihminen nähdään.

Tekoälyjärjestelmien päätöksentekoa ei myöskään ohjaa logiikka. Tämä johtuu algoritmisen päätöksenteon luonteesta, joka pohjaa korrelaatioihin syy-yhteyden sijaan.<sup>386</sup> Niiden käytössä ongelmaksi muodostuu päätöksenteon selitettävyyden ja rationaalisuusvaje, joka on yksi niistä tekoälyn ominaispiirteistä, joilla sääntelytarve valmisteluasiakirjoissa perusteltiin.<sup>387</sup> Myös kuten HCI-tutkimuksen puolella esitettiin, läpinäkyvämyys on juuri korkean autonomian asteen järjestelmille ominaista. Läpinäkyvämyydellä tarkoitettiin, ettei syötteen ja lopputuloksen välinen yhteys välttämättä ole selvä. Oikeudellisen ratkaisutoiminnan kannalta on

---

381 Smith 1997 s. 313.

382 Koulu 2020 s. 720-735.

383 Koulu 2018 s. 851.

384 Tekoälynsäädöksen johdanto-osan perustelukappale 12.

385 Ks. Koulu 2018 s. 855-875.

386 Esim. Malgieri 2019 s. 187.

387 Euroopan komissio COM(2020) 65 final s. 11.

hyvin ongelmallista, jos päätökseen johtaneita perusteluja ei voida esittää asianmukaisella ja ymmärrettävällä tavalla.<sup>388</sup>

Tekoällysäädöksessä ei juurikaan käsitellä niitä syitä, joiden vuoksi tekoälyjärjestelmät katsotaan ominaisuuksiensa puolesta sopimattomaksi oikeudelliseen päätöksentekoon. Myös oikeudellisen päätöksenteon kohdalla valitut lähestymistavat eli ihmistoimijuus ja ihmisen päätöksenteko määrittävät, kuinka aihetta käsitellään. Koska ihmisen päätöksenteko on yksi säädöksen peruseriaa-asteista, ei järjestelmän ominaisuudet ihmisen mahdollisena korvaajana tule arvioiduksi. Säädöksen vaatimusten kautta järjestelmien heikkoudet päätöksentekijöinä voidaan toki päätellä.<sup>389</sup> Suurimpana sanomattomana haasteena oikeudelliselle tai oikeudellisia vaikutuksia saavalle päätöksenteolle vaikuttaa säädöksessä muodostuvan luottamuksen puute tekoälyjärjestelmiä ja niiden puolueettomuutta kohtaan. Tämä ilmenee parhaiten siitä, että jo voimassa olevat syrjimättömyyttä koskevat kiellot toistetaan tekoällysäädöksen lainkäyttöä koskevissa konteksteissa.<sup>390</sup> Kuitenkin eksplisiittiseksi haasteeksi mainitaan vaikeus tarkoituksenmukaiseen tiedonsaantiin tekoälyjärjestelmän toiminnoista ja siitä johtuva vaikeus haastaa tuloksia oikeudessa eli käyttää puolustus-oikeutta.<sup>391</sup> Tämä vastaa oikeuskirjallisuudessa tunnistettuja ongelmia järjestelmien päätöksenteossa.

Oikeudellisen päätöksenteon monimutkaisuuden korostaminen on yhä vallalla oleva diskurssi. Ihmisen sanotaan olevan ylivoimainen monimutkaisessa päätöksenteossa ja rutiiniluontoiset päätökset voidaan jättää järjestelmille.<sup>392</sup> Ihmisen osallisuutta edellytetään etenkin sellaisessa päätöksenteossa, jossa on kyse moninaisten sosiaalisten ja yhteyskunnallisten normistojen soveltamisesta.<sup>393</sup> Näin ollen vaikuttaa siltä, että oikeusinformatiikkojen korostama ihmisen päättelyn inhimillisintuutiivinen ulottuvuus on korvaamatonta. Inhimillisen päättelyn katsotaan tuovan päätöksille legitimitettiä.<sup>394</sup> Inhimillisestä ulottuvuudesta ei olla valmiita luopumaan, vaikka nykyiset järjestelmät oppisivat nimenomaan tämän inhimillisen ajattelun pohjalta.

---

<sup>388</sup> Koulu 2018 s. 857.

<sup>389</sup> Esim. selitettävyys, läpinäkymättömyys, eheys. Nämä ovat säädöksen valmisteluasiakirjoissa tunnistettuja tekoälyjärjestelmien erityispiirteitä.

<sup>390</sup> Tekoällysäädös johdanto-osan perustelukappale 42. Tätä voimassa olevien kieltojen toistamista kuvattiin jakson alussa teknologiasääntelyn haasteita käsittelevässä kohdassa.

<sup>391</sup> Tekoällysäädös johdanto-osan perustelukappale 59.

<sup>392</sup> Ks. esim. HE 145/2022 vp s. 66; Valtiovarainministeriö 2023 s. 22-23.

<sup>393</sup> Enqvist 2023 s. 511.

<sup>394</sup> Koulu 2020 s. 734.

Sen sijaan, että keskustelussa keskityttäisiin tekoälyjärjestelmien ominaispiirteiden aiheuttamiin legitimizeettiongelmiin oikeudellisessa päätöksenteossa - esimerkiksi selitettävyysvajeesta aiheutuvaan päätöksen perusteluiden puutteeseen tai koulutusdatan syrjivien piirteiden kerrannaisvaikutuksiin - palataan oikeudellisessa keskustelussa osittain korostamaan ihmisen ja järjestelmän heikkouksia ja vahvuuksia Fittsin listauksen tapaan. Keskustelu on omiaan ylläpitämään käsitystä dikotomisesta vuorovaikutussuhteesta, eikä sen avulla päästä käsittelemään ydinkysymyksiä inhimillisen päätöksenteon säilyttämisen taustalla.

#### 5.1.4 Lopuksi

Oikeudellinen ihmisen ja järjestelmän vuorovaikutuskäsitys on tekoälysäädöksen tavoitteiden ja niiden toteuttamiseksi valitun keinovalikoiman mukainen. Jaksossa osoitin sekä oikeudellisen vuorovaikutuskäsityksen että tekoälysäädöksen olevan sisäisesti jännitteinen. Jännitteisyys johtuu ennen kaikkea oikeuden antroposentrismistä, joka on yhteinen nimittäjä kaikille edellä käsitellyille oikeudellisen vuorovaikutuskäsityksen muodostaville vaatimuksille eli ihmisvalvonnalle, ihmisen toimijuudelle ja ihmisen suorittamalle päätöksenteolle. Oikeuden antroposentrismi on haastava lähtökohta digitalisoituvan oikeuden aikakaudella, jossa järjestelmien toimijuus tosiasiallisesti osittain tai kokonaan korvaa ihmisen toimijuutta tai ainakin tulee osaksi ihmistoimijuutta.

Jakson aloituskappaleessa luetellut tasot, joilla teknologia ja sen sääntely aiheuttaa yhteiskunnalle ja oikeudelle haasteita, on kaikki tunnistettavissa tekoälysäädöksestä. Ensinnäkin järjestelmien toimijuuden tunnistetaan muokkaavan ihmistoimijuutta ja haastavan siten yhteiskuntamme perimmäisiä arvoja, joille oikeusjärjestelmä rakentuu.<sup>395</sup> Esimerkiksi teknologian mahdollistama massavalvonta uhkaa kokemusta yksityisyydestä ja saattaa vaikuttaa rajoittavasti demokraattiselle yhteiskunnalle keskeisten oikeuksien kuten kokoontumisvapauden käyttämiseen. Toiseksi järjestelmille säädöksessä tunnustetut ominaisuudet kuten reaktiivisuus, responsiivisuus ja vaikutukset ympäristöön pakottavat harkitsemaan niiden toimijuutta, vaikkei järjestelmät ole perimmäisesti arvokkaita tai "vapaita" oikeuden tarkoittamalla tavalla. Teknologian kehitys aiheuttaa oikeusjärjestelmämme

---

<sup>395</sup> Keskustelussa on nostettu esiin ajatus uudistettavasta yhteiskuntasopimuksesta. Kysymyksen keskiössä on, halutaanko teknologian muuttamat arvot hyväksyä. Ks. Kyriakou & Otterbacher 2023.

keskeisten käsitteiden kuten oikeussubjektiuden ja siihen liittyvien toimijuuden kriteerien uudelleen arviointia. Viimeisenä niin sanottuna pintatason haasteena on tarkoituksenmukaisen sääntelyn luominen. Tekoälysäädöksessä omaksutuista keinoista etenkin ihmisvalvonta on todettu tehottomaksi ja siten sääntelyn tuoma suoja voidaan kyseenalaistaa tältä osin.

Tekoälysäädöksessä ihminen pysyy sääntelyn kohteena ja järjestelmien toimijuus koitetaan palauttaa ihmiseen. Ihmisen ja järjestelmän muodostamaa toimijaa ei nähdä omana erityislaatuisena kokonaisuutena, joka on enemmän kuin osiensa summa ja näin ollen vaatii omanlaista sääntelyä, jotta oikeusvarmuus ja sääntelyn legitimitetti voidaan säilyttää. Ihmisen ja järjestelmän vuorovaikutuksen osalta tuleva sääntely ei vaikuta tarkoituksenmukaiselta ottaen huomioon HCI-tutkimuksen tulokset. Päinvastoin dikotominen vuorovaikutuskäsitys on vanhanaikainen ja sopii huonosti yhteen tekoälyjärjestelmien kanssa vuorovaikuttamiseen. Huomionarvoista on, että järjestelmät kehittyvät koko ajan ja siten oikeudellinen vuorovaikutuskäsitys tulee ajan kuluessa olemaan yhä vain kauempana HCI-tutkimuksen ihanteista.

## 6 Loppupäätelmät

Tekoälysäädöksessä omaksuttu tapa hahmottaa vuorovaikutusta noudattaa automaation (alku-)aikakaudella syntyneitä HCI-malleja. Teknologian kehitys on johtanut yhteistyöulottuvuuden korostamiseen vuorovaikutuksessa ja ihmisen ja järjestelmän symbioottinen suhde ei ole enää tiedefiktiota vaan tunnustettu HCI-tutkimuksessa tavoitteeksi erilaisia ihminen-teknologia malleja suunniteltaessa. Tuleva lainsäädäntö ei valitettavasti pysty vastaamaan tähän tavoitteeseen. Osa oikeusjärjestelmästä johtuvista syistä on eritelty edellä.

Yhtenä epäonnistumisen syynä voi olla tieteiden välisen vuoropuhelun haasteet. Käsitteiden merkityssisältö on muuttunut jossain vaiheessa HCI-tutkimuksen tulosten muuttuessa ensin järjestelmäsuunnittelua koskeviksi ohjeiksi ja standardeiksi, sitten oikeudelliseen kontekstiin sääntelyä valmisteleviksi politiikkadokumenteiksi ja lopulta pian voimaan tulevaksi lainsäädännöksi. Tarkoitin, että suhteessa muihin tieteisiin oikeus ymmärtää ihmisen merkityksen vuorovaikutuksessa hyvin kirjaimellisesti omasta antroposentrisestä näkökulmastaan. Siinä missä järjestelmäsuunnittelijoille ihmiskeskeisyys on suunnittelun lähtökohta, jota voidaan käyttää myös korkean yhteistyön järjestelmiä luotaessa, oikeus muuttaa tämän auktoriteettiasemaksi. Oikeudellinen ihmiskeskeisyys edellyttää ihmisvalvontaa ja ihmiseen palautettavaa päätöksentekoa. Vuoropuhelun kannalta ongelmatonta ei ole myöskään se, että valmisteluaineistosta alkaen oletettavia keinovalikoiman taustalla ei tuoda esiin. Oikeus ei siten ole turhan itsetietoinen rajoitteista, joita se asettaa keinovalikoimalle oikeudelliseen kontekstiin siirryttäessä.

Vaikka oikeudellinen vuorovaikutus on jämähtänyt automaation ajan mallinnuksiin, on syytä muistaa tekoälysäädöksen riskiperusteinen lähestymistapa. Moni tekoälyjärjestelmä jää siten säädöksen määräyksien ja sääntelystä nousevien vuorovaikutusrajoitteiden ulkopuolelle. Näiden järjestelmien ja ihmisten välinen vuorovaikutus voidaan vapaasti rakentaa vastaamaan symbioosin ihannetta, kuten vaikuttaakin olevan trendi HCI-alalla. Lisäksi tekoälysäädöksen kohdat, jotka kuvaavat matalan riskin ja korkean riskin rajavetoa antavat viitteitä siitä, että yhteistyömäinen vuorovaikutus hyväksytään ihmisen ja matalan riskin järjestelmien välille oikeudellisestakin näkökulmasta. Dikotomia omaksutaan siinä vaiheessa, kun oikeushyvät käyvät arvokkaiksi ja muuta kuin ihmisjohtoista vuorovaikutuksen

hahmotustapaa ei oikeusjärjestelmän ja poliittisen ilmapiirin rajoissa ole helposti hyväksyttävissä.

Poliittisten ja yleisten mielipiteiden merkitystä lähestymistapoja valittaessa ei voi väheksyä. Riskien esittely legitimoii sääntelyn tarpeen, mutta riskien esittämistapa voi aiheuttaa vääristyneistä käsityksiä siitä, miten riskit muodostuvat. Mikäli riskejä ei kytketä järjestelmien suunnittelu- ja käyttötapoihin, ihmisen toimijuus on hankala havaita. Keskusteluissa ei korostu tällöin ihmisten oma rooli muutoksessa, vaan ihmiskunta on ajopuuna teknologian kehityksen viedessä kohti tuntematonta päämäärää. Sääntely on toki selvä osoitus pyrkimyksestä vaikuttaa teknologissosiaaliseen todellisuuteen. Tekoälysäädöksessä kuitenkin toistetaan uskomuksia itsessään riskialttiista teknologiasta ja osin tästä muodostuu säädökseen ristiriitaisuutta.

Tekoälysäädöksen perusteella vaikuttaa selvältä, että järjestelmien toimijuus etäännyy ihmisen toimijuudesta ja järjestelmät muokkaavat ihmisen toimijuutta. Tarve tekoälyjärjestelmien sääntelylle on ilmeinen. Olemassa olevat oikeussäännöt eivät ole riittävällä tavalla sovellettavissa järjestelmien toimijuuden aiheuttamaan laadulliseen muutokseen vuorovaikutussuhteissa. Sääntely tarvitsee uudelleen kohdistusta ja osittain siksi tekoälysäädöksessä kielletään jo kiellettyjä menettelyitä. Onnistunut sääntelyn vaatisi kuitenkin taustalle tutkimustietoa ja aavistuksen enemmän itsetietoisuutta. Jos lainsäädäntöä tehdään paljastamatta keinovalikoiman taustaoletuksia, valittujen tapojen kriittinen arviointi on haastavampaa. Oikeuden antroposentrismien kannalta tämä tarkoittaisi esimerkiksi sen itsestäänselvyden toteamista, että hierarkia suhteessa järjestelmään täytyy säilyttää, koska lainsäädäntö ei tunnista muuta toimijaa. Tämä toteamus toisi näkyväksi haasteen, jonka teknologia aiheuttaa oikeudelle. Siten keskustelu sääntelymalleista voisi helpottua ja kohdistuisi muutostarpeeseen, joka on mahdollisesti laajempi, kuin yksittäinen teknologiaspesifi säädös.