



**LAU
REA**

AMMATTIKORKEAKOULU
University of Applied Sciences

LAUREA LONG | 2/2023

Asko Mononen, Ari Alamäki ja Olli Laintila

**Vastuullinen tekoäly –
yriytysten yhteiskunnallinen lupaus
tekoälyn soveltamiseen**

Abstrakti

Tekoälyn etiikka (Ethics of AI) ja luotettava tekoäly (Trustworthy AI) ovat olleet jo vuosia keskeisiä tutkimus- ja kehitysteemoja tekoälytutkimuksessa. Viime vuosina varsinkin suuret yritykset ovat nostaneet keskiöön myös vastuullisen tekoälyn käsitteen (Responsible AI) mutta konkreettisia tekoja tarvitaan lisää. Vastuullisella tekoälyllä halutaan korostaa yritysten yhteiskunnallista ja sosiaalista vastuuta kehitettäessä ja myytäessä tekoälyratkaisuja. Valtionhallinto, koulutus ja terveydenhuolto ovat tekoälyn kehittäjille suuria markkinoita, joten vastuullisen tekoälyn käsitteen lanseeraaminen osana markkinointia on ymmärrettävää. Tässä artikkelissa pohdimme, mikä on vastuullisen tekoälyn suhde eettiseen ja luotettavaan tekoälyyn ja mitä vastuullisuus tekoälyssä voisi tarkoittaa käytännössä.

Asiasanat: vastuullinen tekoäly, tekoälyn etiikka, luotettava tekoäly, AI, tekoäly

Vastuullinen tekoäly – yritysten yhteiskunnallinen lupaus tekoälyn soveltamiseen

MAAILMALLA ON USEITA esimerkkejä siitä, miten sosiaaliset ja yhteiskunnalliset näkökulmat liittyvät tekoälyn toimintaan. Monet yritykset ovat havainneet tekoälyn liittyvät sosiaaliset ja yhteiskunnalliset kysymykset kun tekoälyn tulokset eivät ole vastanneet odotuksia, ne ovat olleet vinoutuneita tai epätarkkoja. Kuitenkin konkreettisia vastuullisuuteen liittyviä ratkaisuja, kuten luottamuksen rakentamista ja avoimuutta kaivataan edelleen lisää (Candelon, et al. 2021). Yritysten kannalta on tärkeää tiedostaa miten tekoälyn vastuullisuus, eettisyys ja luotettavuus suhteutuvat toisiinsa ja kuinka yrityksen tekoälyn käyttöön liittyvää vastuullisuutta voisi kehittää.

Automaattisia päätöksiä on tehty jo vuosia mm. Verohallinnossa ja Kelassa. Kela antaa vuosittain 2 - 3 miljoonaa etuuspäätöstä, joista vuonna 2020 jo yli puolet tehtiin automaattisesti ilman etuuskäsittelijän osallistumista. Ongelmana on ollut virkavastuun puuttuminen koneelta. Asian lakipohja ratkaistiin niin, että kone voi jatkossa ratkaista asiat, jotka eivät vaadi tapauskohtaista harkintaa. Tämä tarkoittaa siten päätöksiä, jotka ovat niin kaavamaisia, että ne voidaan pukea käsittelysäännöiksi (Eduskunta, 2023). Tämä sulkee pois esimerkiksi todennäköisyyksiin perustuvat tekoälymallit.

Yhdysvalloissa ja Kanadassa on tutkittu ammattilaisia ja ei-ammattilaisia ja todettu, että syrjintää esiintyi tekoälyavusteisten mallien kanssa, mutta ei ilman tekoälyä. Esimerkiksi jos hätäkeskuspäivystäjälle mainittiin uhrin etninen tausta ja uskonto, se aiheutti syrjintää. Tutkimuksen mukaan tiettyjen ryhmien soittaessa apua lähetettiin poliisipartio muita useammin. Vastaavia vinoumia on dokumentoitu myös oikeuden ja lääketieteen parissa (Adam et al., 2022). Toisin sanoen lisäämällä ihminen mukaan päätöksentekoon (human-in-the-loop) ei aina saada lisää oikeudenmukaisuutta.

EETTINEN, LUOTETTAVA VAI VASTUULLINEN TEKÖÄLY

Liike-elämässä ja julkisella sektorilla on tehty lukuisia selvityksiä, raportteja ja tutkimuksia tekoälyn etiikasta (ethics of AI), luotettavasta tekoälystä (trustworthy AI) ja vastuullisesta tekoälystä (responsible AI). Monelle lukijalle ja alan toimijalle on näiden kolmen käsitteen suhde ja asema on epäselvä (ks. esim. Väänänen 2021).

Tästä johtuen päätimme kysyä asiaa ChatGPT₄-tekoälyltä, kuten nykyään on tapana tehdä. ChatGPT₄ on laaja kielimalli (LLM), joka on koulutettu laajalla internet-datalla (2021 asti). Se on otettu mukaan jopa tieteellisiin artikkeleihin yhtenä kirjoittajana (ks. esim. Osterrieder & GPT 2023), mikä kertoo siitä, että sen tuottamalle tekstillä annetaan paljon painoarvoa puutteista huolimatta. Kysyimme ChatGPT₄:tä, miten se määrittelee eettisen, luotettavan ja vastuullisen tekoälyn ja niiden suhteen toisiinsa. Saimme seuraavan vastauksen:

“Eettinen tekoäly on suunniteltu noudattamaan eettisiä periaatteita ja arvoja, luotettava tekoäly tuottaa ennakoitavia ja virheettömiä tuloksia, ja vastuullinen tekoäly ottaa huomioon tekoälyn käytön vaikutukset yhteiskuntaan ja ympäristöön. Nämä käsitteet liittyvät toisiinsa, sillä eettisyys, luotettavuus ja vastuullisuus ovat tärkeitä tekijöitä tekoälyn kehittämisessä, ja ne vaikuttavat toisiinsa. Tavoitteena on kehittää tekoälyä, joka on sekä eettinen, luotettava että vastuullinen.”

ChatGPT4:n vastaus on linjassa tämän artikkelin kirjoittajien kokemuksen ja tieteellisen kirjallisuuden näkemysten kanssa. Vastuullinen tekoäly keskittyy hyvin pitkälle yritysten yhteiskunnalliseen ja sosiaaliseen vastuullisuuteen. Eettinen tekoäly puolestaan on erittäin tärkeä viitekehys esimerkiksi tekoälyn soveltamisessa koulutukseen ja terveydenhuoltoon. Luotettava tekoäly -käsitteen alla tehdyt kehityshankkeet keskittyvät enimmäkseen tekoälyn ja datan teknisiin ominaisuuksiin, ja niiden tavoitteena on, että tekoäly tuottaa teknisesti luotettavia, laadukkaita ja reiluja tuloksia.

Määritelmät vastuullisesta tekoälystä hakevat vielä muotoaan. Accenturen (2023) mukaan vastuullinen tekoäly on tekoälyn kehittämistä ja soveltamista hyvässä tarkoituksessa voimaannuttamaan ihmisiä ja liiketoimintoja niin, että sillä on reilu vaikutus asiakkaisiin ja yhteiskuntaan. Yleisesti vastuullinen tekoäly huomio tekoälyn sosiaalisen, yhteiskunnallisen ja kestäväan kehitykseen liittyvän vaikutuksen. Näin sen soveltajien ja kehittäjien tulee myös tiedostaa oma vastuunsa tekoälyn suhteen.

VASTUULLISUUS NOUSUSSA TEKÖÄLYKESKUSTELUSSA

Tekoälyyn on liitetty useita sosiaalisia ja yhteiskunnallisia haasteita. Osa tutkijoista kritisoi puutteellisesta ope-
tusdatasta johtuvia tekoälymallien vinoumia, ohjelmoijien homogeenista joukkoa sekä automaattisia päätöksiä. Digitaalisten palveluiden algoritmit aiheuttavat sosiaalisia kuplia koska ihmiset näkevät erilaisissa sovelluksissa ja alustoilla vain heidän profiiliinsa liitetyt sisällöt. Samoin ihmiset usein näkevät pelkästään heidän profiiliaan tukevia sisältöjä varsinkin sosiaalisessa mediassa ja kauppapaikoilla.

Myös yksityisyyteen liittyvät haasteet ovat osa yhteiskunnallisia haasteita. Verkossa tapahtuva tiedonkeruu ja sen analysointi ovat jo pitkään huolestuttaneet kuluttajia ja päättäjiä. Euroopan unioni on reagoinut tähän tietosuoja-asetuksella ja kohta julkaistavalla AI Act:llä, joka tuo tekoälyn soveltamiseen säädöksiä mm. kasvojen tunnistukseen, sosiaaliseen luokitteluun ja muihin soveltamiskohteisiin.

Tekoälyn soveltaminen tuo merkittävästi enemmän vastuuta yrityksille kuin perinteisen tietotekniikan soveltaminen. Erityisesti ChatGPT ja pian myös Google Bard AI:n kyky keskustella ihmismäisesti ja tuottaa uskottavan oloista sisältöä ovat hämmästyttäneet ihmisiä kautta maailman. Verkossa on yhä enemmän valeutisia ja disinformaatiota, jota on usein vaikea erottaa totuudesta. Lisäksi verkossa on paljon kaupallista ja poliittista mainontaa ja kuluttajien käyttäytymistä ohjaavaa vaikuttamista, jota on vaikeaa tai mahdotonta erottaa objektiivisesti tuotetusta tutkitusta tiedosta. Tekoäly pystyy suuntaamaan vaikuttamista entistä tarkemmin käyttäjäprofiiloinnin kautta. Valtiot ja kulttuurit ja niiden lainsäädäntö ja käytänteet vaikuttavat myös vahvasti tekoälyn vastuullisuuteen.

VOIKOTEKOÄLY OLLA ITSESSÄÄN ARVOVAPAATA?

Keskusteltaessa tekoälyn vastuullisuudesta, tulee ensin tarkastella voiko teknologia ylipäättään olla arvovapaata. Keskeinen kysymys teknologian suhteen onkin pitkään ollut se, onko teknologia arvoneutraalia vai ei (Du & Xie 2021). Aikaisemmin oletettiin, että teknologia on työkalu ilman varsinaisia eettisiä tai vastuullisuuteen liittyviä ominaisuuksia kuten on kynät, vasarat ja kirveet. Käytännössä on havaittu, että esimerkiksi opetusteknologian käyttö on saattanut heikentää joidenkin opettajien opetusta ja siten oppilaiden oppimistuloksia (Young 2004). Oppilaat eivät useinkaan voi vaikuttaa siihen mitä opetusteknologiaa käytetään kouluissa, joten opetusteknologioiden käyttöön voi liittyä merkittäviä sosiaalisesti eriarvoistavia haasteita.

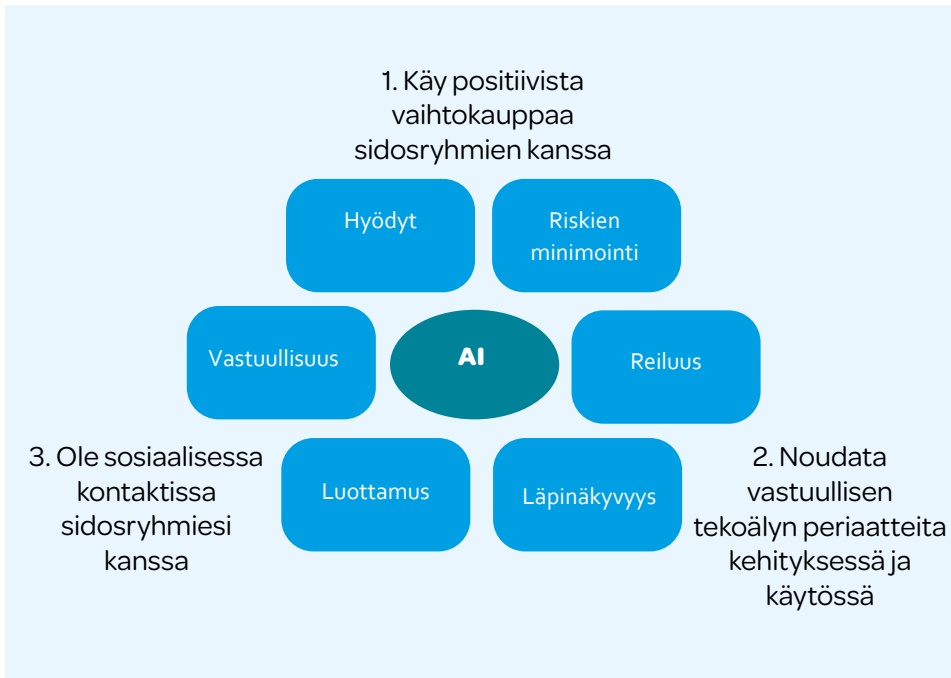
Tekoäly on muuttanut tätä asetelmaa merkittävästi. Tekoäly kykenee ohjaamaan, suosittelemaan ja jopa tekemään itsenäisesti päätöksiä, joten sitä on vaikea pitää arvovapaana. Autonomia on eräs eettisen tekoälyn peruseriaatteita, jolloin käyttäjillä tulisi olla mahdollisuus valita toinen apuväline, jos kyseinen tekoälypohjainen teknologia haittaa esimerkiksi oppimista oppilaan oppimis- tai hahmottamisvaikeuksien takia. Tekoälyllä on siten väärinkäytettynä negatiivisia vaikutuksia.

Nykyään tutkijat kallistuvat sille kannalle, että mikään teknologia ei ole arvoneutraalia edes itsessään. Erityisesti tekoälyn kohdalla yleinen käsitys on se, että tekoäly voi jo itsessään fasilitoida ja ohjata käyttäjiä toimimaan epäeettisesti tai vastuuttomasti. Tekoäly ja monet muutkin teknologiat vaikuttavat, miten ihmiset näkevät ja havainnoivat maailmaa sekä ne ohjaavat käyttäjiä toimimaan tietyllä tavalla ja vaikuttavat ihmisten väliseen vuorovaikutukseen ja sen dynamiikkaan (ks. Esim. Dirin, et al. 2019; Du & Xie 2021; Roberts & David 2016).

Vastuu on kuitenkin monimutkaisempi asia kuin tekoälyn sisältämä arvolutaus. Nykyiset tekoälyratkaisut ovat monimutkaisia kokonaisuuksia, joiden eri osat ja data voivat sijaita jopa eri mantereella pilvipalveluteknologian ansiosta. Perinteisesti vastuun oletetaan olevan sillä taholla, joka tarjoaa tekoälypohjaisen ratkaisun käyttäjille, kehittää ratkaisun tai omistaa sen. Harvemmin vastuuta sysätään loppukäyttäjille, joiden näkökulmasta tekoäly-sovellukset ovat usein "mustia laatikoita" joiden sisäänrakennettua toimintaa on vaikea nähdä ja ymmärtää.

MALLITEKOÄLYN VASTUULLISUUDEN KEHITTÄMISEEN YRITYKSISSÄ

Aivan kuten minkä tahansa ihmisen toimintaan vaikuttavan teknologian kehityksessä, myös tekoälyn kehittämisessä pitää ottaa huomioon sosio-teknisiä kysymyksiä. Candelon et al. (2021) ovat kehittäneet tekoälyn vastuullisuuden jäsentämiseen ja kehittämiseen "Sosiaalinen lisenssi tekoälylle"-mallin (kuvio 1). He toteavat, että yritysten tulee noudattaa vastuullisen tekoälyn kehittämisen periaatteita kehitystyössä, varmistaa että sidosryhmät hyötyvät tekoälystä enemmän kuin se aiheuttaa kustannuksia ja yritysten tulee havainnollistaa, että heihin voi luottaa ja he tulevat toimimaan vastuullisesti.



Kuvio 1. Sosiaalinen lisenssi tekoälylle-malli (Candelon et al. 2021).

JOHTOPÄÄTÖKSET

Vastuullisuuden kysymykset liittyvät yrityksissä pitkälti tekoälyn keskeneräisyydestä johtuvien puutteiden hallintaan, avoimuuteen ja luottamuksen rakentamiseen. Mika Pantzar (Hupsis, tein virheen -radio-ohjelma 2023) vertaa tekoälyä vielä keskenkasvuiseen lapseen, jota ei pitäisi ennen aikaisesti liikaa kritisoida. Onneksi meillä ihmisillä on mahdollisuus vaikuttaa vastuullisen tekoälyn kehitykseen.

Nykyinen tekoäly on vielä kaukana yleisestä älykkyydestä ja tekoälyn älykkyyteen liittyy vielä paljon rajoittavia asioita ihmiseen verrattuna. Nykyiset tekoälysovellukset eivät kykene vielä intuitiiviseen psykologiaan tai fysiikkaan, kuten ennakoimaan tai päättämään intuitiivisesti asioita riittävällä tarkkuudella, eivätkä ne osaa ihmisen tavoin siirtää oppimaansa helposti kontekstista toiseen. Tekoälyltä puuttuu tilannetaju ja emotionaalinen älykkyyden toimiessaan monimutkaisissa ympäristöissä ja tekoälyn todellista uuden luomiskykyä on myös kritisoitu (ks. esim. Lake, et al., 2017; Mogensen, 2023). Kehitys menee kuitenkin suurin harppauksin eteenpäin varsinkin erikoistuneissa tekoälysovelluksissa. Tämän takia tekoälyn soveltamisessa korostetaan psykologista ja sosiaalista ymmärrystä (Zhao, 2018; Mariani, Perez-Vega & Wirtz 2022).

Nyt on oikea aika kiinnittää yrityksissä huomiota tekoälyn vastuullisuuteen. Konkreettisimmin se tapahtuu vastuullisen tekoälyn kehittämisen ja hyödyntämisen periaatteiden soveltamisen kautta omiin käytäntöihin. Lisäksi vastuullinen tekoälyn kehittäminen ja hyödyntäminen edellyttää sidosryhmäyhteistyön kehittämistä ja vuorovaikutuksen lisäämistä aiheen tiimoilta. Näin tekoälyn kehitys tapahtuu käyttäjakeskeisemmin huomioiden sidosryhmien intressit.

Lähteet

Accenture. 2023. Responsible AI: Scale AI with confidence. Luettavissa:

<https://www.accenture.com/us-en/services/applied-intelligence/ai-ethics-governance>

Adam, H., Balagopalan, A., Alsentzer, E. et al. 2022. Mitigating the impact of biased artificial intelligence in emergency decision-making. *Communications Medicine* 2, 149. <https://doi.org/10.1038/s43856-022-00214-4>

Candelon, F., di Carlo, R. C. & Mills, S. D. 2021. AI-at-Scale Hinges on Gaining a 'Social License'. *MIT Sloan Management Review*, 63(1), 1-4.

Dirin, A., Alamäki, A. & Suomala, J. 2019. Digital amnesia and personal dependency in smart devices: a challenge for AI. *Proceedings of Fake Intelligence Online Summit 2019*.

Du, S. & Xie, C. 2021. Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities. *Journal of Business Research*, 129, 961-974.

Eduskunta. 2023. Yleislainsäädäntö automaattiselle päätöksenteolle. Viitattu 19.3.2023 Lainattu 19.3.2023

https://www.eduskunta.fi/FI/naineduskuntatoimii/kirjasto/aineistot/kotimainen_oikeus/LATI/Sivut/yleislainsaadanto-automaattiselle-paatoksenteolle.aspx

Lake, B. M., Ullman, T. D., Tenenbaum, J. B. & Gershman, S. J. 2017. Building machines that learn and think like people. *Behavioral and brain sciences*, 40, e253.

Mariani, M. M., Perez Vega, R. & Wirtz, J. 2022. AI in marketing, consumer research and psychology: A systematic literature review and research agenda. *Psychology & Marketing*, 39(4), 755-776.

Mogensen, K. 2023. Artificial Intelligence a Myth? Exploring whether human sense-making can be surpassed by AI. *FARSIGHT futures reviewed*. <https://farsight.cifs.dk/is-artificial-intelligence-a-myth/>

Osterrieder, J. & GPT, C. 2023. A Primer on Deep Reinforcement Learning for Finance. Available at SSRN 4316650. <http://dx.doi.org/10.2139/ssrn.4316650>

Roberts, J. A., & David, M. E. 2016. My life has become a major distraction from my cell phone: Partner phubbing and relationship satisfaction among romantic partners. *Computers in human behavior*, 54, 134-141.

Väänänen, K. 2021. Tekoäly kaupunkikehityksessä – ihmiskeskeisen suunnittelun näkökulma Luettavissa:

<https://www.eduskunta.fi/FI/vaski/JulkaisuMetatieto/Documents/EDK-2021-AK-405736.pdf>

Hupsis, tein virheen! Lista turvallisuusuhkista puutteellinen? Tekoälyä täytyy saada kritisoida. Radio-ohjelma 15.2.2023. Yle Areena –verkkopalvelu. Kuunneltavissa: <https://areena.yle.fi/podcastit/1-64892274>

Young, J. 2004. When good technology means bad teaching: Giving professors gadgets without training can do more harm than good in the classroom, students say. *The chronicle of higher education*, 51(12), A31-A37.

Zhao, W. W. .2018. Improving social responsibility of artificial intelligence by using ISO 26000. In *Iop conference series: Materials science and engineering*. Vol. 428, No. 1, p. 012049. IOP Publishing.

Copyright® tekijät ja Laurea-ammattikorkeakoulu 2023

Kannen kuva: Deep Mind on Unsplash

Tekijät:

Asko Mononen toimii lehtorina Laurean Digital Living Labissa ja 3AMK tekoälytiimissä.

Ari Alamäki toimii yliopettajana Haaga-Heliassa ja 3AMK tekoälytiimissä

Olli Laintila toimii lehtorina Haaga-Heliassa

CC BY-SA 4.0

ISSN 2954-2170