R. Linturi Oyj
Risto Linturi

LAUSUNTO

14.9.2021

Asia: U 28/2021 vp Valtioneuvoston kirjelmä eduskunnalle komission ehdotuksesta Euroopan parlamentin ja neuvoston asetukseksi tekoälyn harmonisoiduksi sääntelyksi (Artificial Intelligence Act). Lausunnossa huomioitu myös E 62/2021 vp Valtioneuvoston selvitys: Komission tiedonanto: Tiedonanto eurooppalaisesta lähestymistavasta tekoälyyn

Aihe: Tekoälyn kansalais-/tukiälynäkökulman huomiointi tekoälyn sääntelyssä

Tausta: Tekoäly kehittyy yhä nopeammin. Kesäkuun uutisina on kerrottu mm., miten Googlen tekoäly kehittää itselleen soveltuvaa piiriteknologiaa paljon ihmisiä nopeammin. Kiinassa on julkaistu samanaikaisesti mm. keskusteluun, musiikkiin ja kuvien tuottamiseen kykenevä, parametrimäärältään maailman suurin tekoäly. DeepMind -tutkijoiden julkaisema tieteellinen artikkeli osoittaa yleisen älykkyyden olevan saavutettavissa tekoälyjärjestelmien laajasti käyttämän vahvistusoppimisen avulla. Omassa käytössäni, oppimateriaalien ja esitysten tuottamisessa käytän rutiinimaisesti useita tekoälysovelluksia ja keskustelen tukiälyn kanssa kohdeyleisöjen tarpeista ennen esitysteni valmistelua. Muutos lupauksista arjeksi on ollut nopea ja vielä tällä vuosikymmenellä on ennakoitavissa radikaali murros sekä arjessa että työelämässä.

Tuoreimpana muutoksena kehittyneimmän yleisön saatavilla olleen keskustelevan tekoälyn GPT-3:n kehittäjä OpenAI on estänyt tavallisilta käyttäjiltä mahdollisuuden personoida tekoäly omiin tarkoituksiinsa parhaiten sopivaksi vedoten eettisiin ongelmiin. Kehitys on tasa-arvon näkökulmasta hyvin ongelmallinen.

https://www.theverge.com/2021/6/10/22527476/google-machine-learning-chip-design-tpu-floorplanning

https://bdtechtalks.com/2021/06/07/deepmind-artificial-intelligence-reward-maximization/

https://reason.com/2021/06/11/amazing-new-chinese-a-i-powered-language-model-wu-dao-2-0-unveiled/

https://www.theregister.com/2021/09/08/project_december_openai_gpt_3/

Tulevaisuusvaliokunnan julkaisemassa raportissa, ”Suomen sata uutta mahdollisuutta 2018-2037”, analysoidaan tekoälyn eri tekniikoita ja sovellusalueita. Raportissa listaamme yli 300 tekoälyn erillistä sovellusaluetta, joiden mahdollinen yhteiskunnallinen vaikutus on arvioitu arvonluontiverkostoittain ja vaadittavat teknologiat kuvattu satojen lähdeviitteiden tuella.

Toisessa samoja havaintoja käsittelevässä, tuoreessa raportissa, ”Kohti parempaa tulevaisuutta”, kiteytetään arvonluontiverkostoittain eri teknologioiden kehityksen mahdollistamat keskeisimmät sosiotekniset murrokset ja arvioidaan näiden yhteiskunnalliset vaikutukset ja jännitteet YK:n kestävän kehityksen Agenda2030 -kriteerien valossa. Tekoälyn ja erityisesti joukkoistetun tukiälyn muodossa käytetyn tekoälyn mahdollisuudet maailman visaisten ongelmien selättämiseen ovat raportin yksityiskohtaisen analyysin mukaan aivan keskeiset.

https://www.eduskunta.fi/FI/naineduskuntatoimii/julkaisut/Documents/tuvj_1+2018.pdf

https://www.eduskunta.fi/FI/valiokunnat/tulevaisuusvaliokunta/julkaisut/Sivut/kohti-parempaa-tulevaisuutta.aspx

Taustaksi mainitsen vielä tulevaisuusvaliokunnalle laatimani lausunnon, jossa on kuvattu lyhyesti tekoälyn keskeisimmät sovellusalueet ja kuvattu erityisesti tukiälyksi soveltuvaa keskustelevaa teknologiaa. Kokeilun tärkein tulos oli osoitus siitä, miten yleinen kielimalli GPT-3 kykenee käyttäjän ohjaamana vastaamaan kysymyksiin kuten luontoaktivisti tai, kuten teollisuuspatruuna. Näissä malleissa keskusteleva ihminen siis vaikuttaa itse keskustelun motiiviin ja tekoälyn asenteisiin samalla tavalla kuin ihminen valitsee sen, keiden toimittajien artikkeleita lukee ja kenen tulkitsemana uutisvirtaa seuraa. Tällä on olennainen merkityksensä sääntelyä pohdittaessa. Tämä näkökulma on tuore ja yleisissä arvioissa käsittelemätön. Kokeilu on kuitenkin jo ehtinyt saada tutkijoiden huomiota.

https://www.eduskunta.fi/FI/vaski/JulkaisuMetatieto/Documents/EDK-2021-AK-369260.pdf

En kirjaa seuraavassa uudestaan edellä kuvatuissa tai valtioneuvoston kirjelmässä esitettyjä yleisiä periaatteita tai perusteluita vaan paneudun valtioneuvoston kirjelmässä esitettyjen asioiden kommentointiin sekä joidenkin sivuhuomioiden tasolla Työ- ja elinkeinoministeriön Perusmuistion (TEM2021-00327) mukaisiin Suomen kannanottoihin.

Yleinen kirjelmän sisältö ja harmonisoinnin tavoitteenasetanta

Yleisellä tasolla Euroopan parlamentin ja neuvoston esitystä tekoälyn harmonisoiduksi sääntelyksi voidaan pitää perusteltuna ja hyvänä avauksena. Pyrkimys eettiseen ja turvalliseen tekoälyyn on hyvä, ja samalla luokittelu eri kriteerein säädeltyihin tekoälysovelluksiin tarpeen. Samaten valtioneuvoston kirjelmä eduskunnalle on asiantuntevasti valmisteltu vaikkakaan ei huomioi myöhemmin esittämiäni puutteita. Puutteet vaarantavat erityisesti vammaisten ja osaamis- ja esiintymistaidoiltaan sekä sosiaalisilta kyvyiltään heikompiosaisten osallistumismahdollisuuksia, mutta saattavat rajoittaa myös luovaa työtä.

On olennaista erottaa tekoäly perinteisistä algoritmisista tietojärjestelmistä kahdesta keskeisestä syystä. Kun puhutaan koneoppimisesta ja oppiminen voi tapahtua ihmisen tavoin luonnollista maailmaa havainnoimalla, esimerkiksi katsomalla ja kuuntelemalla, on käytännössä mahdotonta tietää, mitkä kaikki havainnot synnyttävät muistijälkiä tekoälyn mahdollisesti satoihin tai tuhansiin miljardeihin parametreihin.

Sanotun vuoksi on helpompi säädellä tekoälyn käyttäytymistä kuin sitä, mitä muistijälkiä tekoälyn parametreihin kertyy. Voidaan toki kuvitella, ettei tekoälylle anneta mitään sellaista tietoa, jonka käytön vaikkapa GDPR kieltää, mutta silloin tekoälyn tulisi olla sokea ja kuuro, ja se tulisi opettaa kategorisoiduilla käsitteillä. Erityistapauksissa tämä on ratkaistavissa, mutta hyvin suuri osa tekoälyn ja erityisesti tukiälyn sovelluksista jäisi saavuttamatta. Sanon tämän, vaikka olen tietoinen, ettei esityksessä nyt irtaudutakaan tekoälyn osalta GDPR:stä.

Tekoälyn osalta regulaatiokehityksessä suuntaudutaan kuitenkin tulkitsemaan järjestelmän toimintaa ja oppimateriaalia parametreihin kertyneen tiedon sijaan. Toinen tärkeä näkökulma on ilmeinen ymmärrys tekoälyn toiminnan hienojakoisuudesta siten, ettei varsinaista päättelyn logiikkaa ole helposti tunnistettavissa, jos päättelyyn vaikuttavia parametreja on miljoonia eikä niillä useinkaan ole käsitteellisiä nimiä. Päätelmien perustelu on parhaimmillaankin siksi saman

kaltaista kuin ihmisillä, jotka tekevät päätöksen ja perusteluita penättäessä keksivät jonkin selityksen, joka kuulostaa rationaaliselta päättelyltä, mutta ei useinkaan ole aukoton tai edes se, miksi päätös todellisuudessa tehtiin. Tämän vuoksi tekoälyn toimintaa tulee valvoa monilla tavoin samoin periaattein kuin ihmistenkin toimia valvotaan.

Pyrkimys sääntelyn harmonisointiin on tärkeä, ja tavoite tekoälyn kannalta suotuisan ympäristön luomiseksi on olennainen, koska EU on selkeästi jäämässä jälkeen sekä USA:n että Kiinan tekoälykehityksen vauhdista. Ilmeistä ei valitettavasti ole, että tämä sääntely riittäisi eron kuromiseen umpeen. Eurooppalaisessa keskustelussa monet voimakkaasti korostetut pelot jättävät monien muiden täysin toisenlaiset pelot ja ihmisoikeudet helposti huomiotta. Vastakkainasettelussa usein korostuu yksityisyyden suoja, mutta jää huomioimatta joukkoistetun teknologian yksilöille tarjoama hajautettu turva sekä robotisaation ja tukiälyn käytön avaama laaja mahdollisuuksien tasa-arvo. Osoitan yksityiskohtia tähän ajatukseen myöhemmin. Jako kiellettyihin käyttökohteisiin, korkean riskin käyttötarkoituksiin, matalan riskin ja minimaalisen tai olemattoman riskin kohteisiin on hyvä, mutta rajankäynnissä on paljon ongelmakohtia juuri näistä näkökulmista katsoen.

Sääntely kohdistuu tulevaisuuteen estäessään järjestelmiä, joita ei vielä ole kehitetty. Siksi vaatii mielikuvitusta, jotta näkisi, mikä kaikki sääntelyn vuoksi jää kehittämättä. Usein lapsi menee tällaisissa hankkeissa pesuveden mukana ja se huomataan, kun jälkikäteen joudutaan luopumaan epäonnistuneista rajoituksista, kun kilpailevat alueet ovat vallanneet omat markkinansa ja osoittaneet uusien käytäntöjen hyödyt ja toisenlaiset tavat haittojen minimointiin. Suojelun kohdistuessa yleisiin uhkiin tulee helposti estettyä asioita, jotka yksilötasolla tarjoaisivat paljon kaivattuja taitoja ja kykyjä, joita lahjakkaimmilla ihmisillä on luonnostaan.

Kokonaisuudessaan tämä näkökulma vaikuttaa sivuutetun ja tekoälyä on käsitelty vanhakantaisen tuottaja-kuluttajanäkökulman avulla, jossa tekoälyä palvelun tuottamiseen käyttää jokin organisaatio ja kansalainen on kuluttajan roolissa. ”Prosumer” tai ”Empowerment” -näkökulma on sivuutettu, jossa tekoäly synnyttää yksilöille mahdollisuuksien tasa-arvoa. Keskityn jatkossa tähän itseäni lähellä olevaan kansalaisnäkökulmaan. Kansalaisia ei pidä suojella tavalla, joka riistää heiltä mahdollisuudet tasavertaiseen osallistumiseen.

Yksittäiset ongelma-alueet tukiälyn näkökulmasta

Seuraavassa on listattu joitakin laajoja ryhmiä koskevia olennaisia yksityiskohtia, jotka tulisi ottaa huomioon säädöksiä tarkennettaessa. Yleinen ajattelu sivuutetaan, koska se on mielestäni terveellä pohjalla ja ongelmat liittyvät lähinnä kriittisiin yksityiskohtiin edellä osoittamistani näkökulmista.

Kielletyt tekoälyn käytännöt

Kielto kohdistuen alitajuisen manipuloinnin käyttöön tavalla, joka voi aiheuttaa materiaalista, fyysistä tai psyykkistä haittaa, on todella ylimalkainen. On huomattava, että tukiälyä voidaan käyttää opetukseen, valmennukseen ja lukuisiin muihin tarkoituksiin, joihin esimerkiksi kirjoja käytetään nyt. Opetettuani manipulointia ja lobbausta ymmärrän alitajuisen manipuloinnin hyvin

laajaksi kokonaisuudeksi, jota käytetään jatkuvasti myynnissä, markkinoinnissa, seurustelussa, politiikassa yms. Haitan aiheuttaminen on myös hämmentävän tulkinnanvarainen asia.

Laveasti tulkittuna tämä kielto estäisi keskustelevan tekoälyn käytön sellaisessa muodossa, jossa itse sitä käytän ja, jossa sitä mm. käytettiin taustoituksessa mainitussa tulevaisuusvaliokunnan suorittamassa kokeilussa. Siis manipulointina voidaan pitää keskusteluapuna, jossa tekoäly auttaa löytämään viitekehyksiä ja argumentointia. Jos ihminen hakee tekoälyltä vastauksia omien harrastustensa, vaikkapa vuorikiipeilyn motivoinniksi ja vuorikiipeilyn riskit realisoituvat, onko vuorikiipeilyn ihannointi ollut haitallista alitajuista manipulointia. Mille tielle ajaudumme, jos kiellämme samoin löysin perustein kirjoja? Ajattelun vaikutukset näkyvät jo OpenAI:n toiminnassa, koska mainittu tulevaisuusvaliokunnan kokeilu erilaisten eettisten periaatteiden vaikutuksesta saman GPT-3 -tekoälyn vastauksissa samoihin kysymyksiin on nyt mahdoton järjestää OpenAI:n kiellettyä valvomattoman ja sensuroimattoman käytön.

Tällaisen regulatiivisen kiellon ulkopuolelle tulisi erityisen selkeästi jättää kaikki sellainen yleinen, käyttäjänsä määriteltävissä oleva tekoäly, joka pyydettäessä esimerkiksi keskustelee ihmisen tavoin, vaikka puhe olisi manipuloivaa, mikäli käyttäjä, joka järjestelmän kanssa keskustelee, on tästä selkeästi tietoinen tai on itse keskustelun luonteen määritellyt. Tällä tavoin laveakaan manipuloinnin tulkinta ei estä ihmisen tavoin keskustelevan tukiälyn käyttöä hänen tarvitessa valmennusapua tehtäviinsä. Näitä tehtäviä, joissa tekoäly voisi auttaa ja, jotka joku voisi manipuloinniksi luokitella on paljon mm. kirkon, kuntoliikunnan, ravitsemuksen, myynnin, opiskelun, taiteen, politiikan ja henkisen valmennuksen piirissä.

Samaten itse nyt rutiininomaisesti käyttämääni puhuvaa apupäätä, joka katsoo kameraan ja on herpaantumattoman vakuuttavan näköinen, vaikka itse luen puheeni paperista katsoen paperia, siinäkin kyse lienee alitajuisesta vaikuttamisesta ja, jos jonkun mielestä puheistani on haittaa, mennään jälleen harmaalle alueelle. Kuvankäsittelyssä on yleisesti suuri määrä alitajuista vaikuttamista, kuten jokainen mainostoimistojen kanssa työskennellyt tietää. Nämä ovat nyt vain esimerkkejä, mutta osoittavat mielestäni, että sekä manipulointi että haitta tulisi määritellä ahtaasti eikä näin yleisellä tasolla. Lisäksi tulisi sulkea esimerkein pois suuri joukko asioita, joita ei tarkoiteta. Näin erityisesti liittyen avoimiin kielimalleihin ja epäsuoraan tekoälyn käyttöön. Siis tilanteisiin, joissa tekoäly auttaa taidoiltaan puutteellisia ihmisiä tuottamaan yhtä vakuuttavia tuloksia kuin mihin taitavat esiintyjät kykenevät. Tätä tasa-arvoistavaa tukiälyn käyttöä ei tulisi säädöksin estää.

Korkean riskin tekoälyjärjestelmät

Rajausta tehtäessä tulee erityisesti huomioida erityisryhmien, kuten näkö- ja kuulovammaisten, muistisairaiden ja muiden toimintakyvyiltään rajoittuneiden ihmisten tarpeet. Pääsääntöisesti tämä tulisi tehdä siten, että vain välttämättömiltä osin nämä saatetaan rajauksen piiriin, koska mekanismin raskaus johtaa helposti siihen, ettei ryhmän tavoitettavaksi riittävän edullisia tuotteita tule markkinoille. Biometristä tunnistusta ei tulisi rajata tavalla, joka estää tavaroita tunnistamasta luvallisia käyttäjiään, tai avustamasta toimintakyvyiltään rajoittuneita ihmisiä tunnistamasta kanssaihmisiään ja kommunikoimasta heidän kanssaan tai selviytymästä itsenäisesti tukiälyn avulla. Tunnistusta ei myöskään tulisi estää tilanteissa, joissa menettelyllä

on olennainen vaikutus sitä käyttävän yksilön omaan turvallisuuteen ja turvallisuuden tunteeseen. Hyväksikäyttötilanteita, joissa käyttäjän oma turvajärjestelmä voisi helposti ehkäistä, on nyky-yhteiskunnassa runsaasti.

”Tiettyjen tekoälyjärjestelmien läpinäkyvyysvelvoitteet koskevat järjestelmiä, jotka ovat vuorovaikutuksessa ihmisten kanssa, joita käytetään havaitsemaan tunteita tai määrittämään yhteys (sosiaalisiin) luokkiin biometristen tietojen perusteella tai joita käytetään tuottamaan tai manipuloimaan sisältöä (”deep fake”). Kun henkilöt ovat tekemisissä tekoälyjärjestelmän kanssa tai henkilön tunteet tai ominaisuudet tunnistetaan automatisoidusti, henkilöille on kerrottava tilanteesta. Jos tekoälyjärjestelmää käytetään sellaisen kuvan, ääni- tai videosisällön tuottamiseen tai käsittelyyn, joka muistuttaa huomattavasti aitoa sisältöä, olisi oltava velvollisuus ilmoittaa, että sisältö on luotu automaattisin keinoin, jollei poikkeukset perustu laillisiin tarkoituksiin kuten lainvalvonta tai sananvapauden toteuttaminen.”

Tässä kohden tulisi varmistaa, että esimerkiksi ihmisten seassa liikkuvat laitteet saavat luokitella ihmisiä voidakseen ennakoida yllättävää käyttäytymistä. Robottiauton on esimerkiksi tärkeää ymmärtää, milloin lapsi saattaa juosta yllättäen maantielle tai, milloin on odotettavissa ihmisten muutoin käyttäytyvän liikenteessä yllättävästi.

Mitä tulee ”deep fake” asiaan, tulisi olla mahdollista korjata omaa vireystilaa tai katseen kohdistusta yms. normaalia tekoälyn avustamaa kuvamanipulointia tästä erikseen vastapuolelle kertomatta. Kaikkien ihmisten puhetaidot ja muut esiintymistaidot eivät ole samalla tasolla, ja tekoälyä tulisi voida tällaisessa käytössä käsitellä proteesin tavoin asiana, joka on yksityinen. Tämän oikeuden toteutuminen tulisi varmistaa. Itse tuotan jo osan omista esiintymisistäni itseni näköisen puhuvan apupään avulla ja minua kiusaisi kyllä, jos joutuisin leimaamaan kaiken, jossa minulla ei ole mahdollisuutta käyttää studiotason järjestelmiä vaan käytän studiossa tuotettua apupäätäni, jonka huulia ja ilmeitä puheeni liikuttaa. Näen asian mahdollisuuksien tasa-arvon kautta.

Globaalin kehityksen näkökulma

Kansallisten etujen huomiointi on olennainen asia liittyen erityisesti Euroopan komission suunnitelmiin tekoälyn edistämiseksi. Työ- ja elinkeinoministeriön perusmuistio keskittyy asiaan organisaatioiden kannalta, mutta kaksi keskeistä näkökulmaa jää vaille mainintaa.

Suomen kielen ja tietopohjan osalta on erittäin olennaista, että EU:n piirissä edistettäisiin standardoituja rajapintoja, joiden avulla pienetkin Euroopan kielet olisivat käyttökelpoisia tekoälysovelluksissa. Laajoja kielimalleja ei tuoteta erikseen kutakin kieltä varten, koska esimerkiksi suomeksi saatava aineisto on liian suppea kielimallin kunnolliseen opetukseen. Siksi opetus on tapahduttava globaalilla aineistolla ja käytön on sujuttava kansallisilla kielillä. Sama koskee kulttuurista ja tiedollista sisältöä. Jopa suomenkielisen puheen laadukas tuottaminen on nyt paljon kapeampialaista, hankalampaa ja kalliimpaa kuin lukuisilla muilla kielillä huolimatta siitä, että olimme selkeä edelläkävijä aiheessa vielä viime vuosituhannella.

Tämän tarkastelun ajattelu lähtee yleisemmästä tukiälyn näkökulmasta, joka vaikuttaa puuttuvan käsittelyssä olevista dokumenteista. Arvioidut dokumentit toimivat hyvin, kun niitä ajatellaan erikoistuneiden organisaatioiden ollessa toimijoita ja kansalaisten ollessa toiminnan kohteita. Ajattelu toimii huonommin, kun tarkastellaan kansalaista aktiivisena geneeristä tekoälyä käyttävänä yksilönä. Dokumentit esiintyvätkin arvomaailmaltaan hallintokeskeisinä ja yrityskeskeisinä kansalaiskeskeisyyden sijaan, ja vaikka kansalaisille pyritään luomaan ajattelun avulla turvallisia palveluita, ei keskitytä siihen, miten kansalaiset itse voivat hyötyä tai jäädä vaille hyötyä tekoälykehityksen edetessä.