LAUSUNTO

Eduskunnan liikenne- ja viestintävaliokunnalle
Asia: VNS 7/2018 vp Valtioneuvoston selonteko tietopolitiikasta ja tekoälystä

Lausunto selonteosta tietopolitiikasta ja tekoälystä

Electronic Frontier Finland - Effi ry kiittää mahdollisuudesta antaa lausunto asiassa. Effi on vuonna 2001 perustettu kansalais- ja asiantuntijajärjestö, joka puolustaa perus- ja ihmisoikeuksia digitaalisessa toimintaympäristössä.

Lähtökohtaisesti Effi on samaa mieltä mietinnön yleisvireen kanssa. Tietopolitiikasta ja tekoälyn käytöstä on hyvä yhteiskunnallista keskustelua niiden yhteiskunnallisen merkittävyyden takia. Lisäksi Suomella on kokoaan suuremmat mahdollisuudet tekoäly- ja tietotekniikkapohjaisilla aloilla ainakin vielä toistaiseksi kun korkeakoulu- ja tutkimusjärjestelmämme on vielä jotenkin kilpailukykyinen. Jatkossa toki koulutukseen ja tutkimukseen on panostettava, jotta maamme asema tiedettäkin harjoittavana sivistysvaltiona säilyy.

Eettisistä perusteista on aitoja erimielisyyksiä

Selonteon kohdassa 1.1 - Mitä on eettinen tietopolitiikka? - todetaan seuraavasti:

"Suomi on vakaa luottamusyhteiskunta. Luottamusta vahvistaa edelleen yhteiskunnan eri toimijoiden sitoutuminen yhteisiin eettisiin periaatteisiin ja arvopohjaan."

Yhteiset eettiset minimiperiaatteet löytyvät ja niihin sitoutuminen määrittyy poliittisen ja muunkin yhteiskunnallisen keskustelun ja toiminnan kautta. Erittelemättömän arvojen kentän näkeminen yhtenäisenä "voimakenttänä" vaatii kuitenkin perustelukseen keskustelua ja asioiden määrittelyä. Vaikka yhteiskunnallisten näkemysten erot ovat Suomessa pienempiä kuin monessa muussa maassa, eroja silti on.

Viime vuoden aikana Effi on useamman kerran lausunut lakialoitteista sellaisia yksityiskohtaisia asioita, jotka palautuvat aitoon erimielisyyteen siitä, mitä voimassaoleva lainsäädäntö ja yleiset oikeusperiaatteet sallivat ja mitä tulisi tehdä ja miksi. Lähellä selonteon tarkoittamaa toimintaa on esimerkiksi hallituksen esitys laiksi sosiaali- ja terveystietojen toissijaisesta käytöstä (HE 159/2017), jolla on tarkoitus mahdollistaa sote-tietojen alkuperäisestä

käyttötarkoituksesta poikkeava, yhteiskuntaa, tutkimusta ja liiketoimintaa hyödyttävä käyttötapa. Effin näkemyksen mukaan esityksessä tehdään tietojen käsittelyn oikeutuksen osalta eräänlainen silmänkääntötemppu: koska arkaluonteisia henkilötietoja ei voi EU:n yleisen tietosuoja-asetuksen mukaan käsitellä näin ja toisaalta tietosuoja-asetuksen mukaan anonymisoitu henkilötieto ei ole henkilötietoa, esityksen ajatus on anonymisoida sote-henkilötiedot, jotta toissijainen käyttö mahdollistuisi. Effin näkemyksen mukaan esityksessä on jätetty kokonaan huomiotta se, että jo tietojen anonymisointi on henkilötietojen käsittelyä, jolle täytyy olla oikeutus. Sellaista ei kuitenkaan ole esitetty ja siten koko henkilötietojen käsittelyn oikeutus on mielestämme virheellinen.

Tämä esimerkki osoittaa, että on olemassa aitoja erimielisyyksiä asioissa, joiden ehkä pitäisi kuulua yhteisten eettisten periaatteiden ja arvopohjan alaan. Koska näin ei selvästikään ole, yhteiskunnallista keskustelua tulee rohkeasti ja väsymättömästi jatkaa, jotta löydetään ne yhteiset minimitekijät, joiden varaan päätökset ja linjaukset rakennetaan.

Käytettyjä käsitteitä on syytä pyrkiä tarkentamaan

Selonteon kohdassa 1.2 - Mitä on tekoäly? - on tekoälyn määritelmä

"Tässä selonteossa tekoälyllä tarkoitetaan tietokoneohjelmia, joiden avulla koneet, laitteet, ohjelmat, järjestelmät ja palvelut voivat toimia älykkäästi, eli joustavasti ja tarkoituksenmukaisesti monimutkaisessa ja osin ennustamattomassa ympäristössä."

Tarkoitus ei liene sanoa että ympäristö on ennustamaton vaan että

1) tekoälyllä pystytään hahmottamaan muuttuvaa tilannetta siten, että
2) tulkinta ei perustu luonnollisella kielellä ilmaistavaan päätelmien joukkoon ja niistä saatavaan lopputulokseen vaan tulkinta on tekoälyjärjestelmän, esim. neuroverkon tila.

On syytä korostaa tekoälyn olennaista luonnetta tässä kohtaa: koska käytetty metodologia ei noudata luonnollisen kielen sääntöjä, toimintaa ja lopputulokseen pääsemisen logiikkaa ei pystytä välttämättä ilmaisemaan osuvasti luonnollisen, ihmisten käyttämän puhutun ja kirjoitetun kielen avulla.

Selonteon sivulla 6 käytetään käsitettä "kapean tekoäly" termi. Alan tutkimuksessa käytetään myös mielestämme osuvampaa ilmaisua "heikko tekoäly" ja sille vastakohtana "vahva tekoäly". Kuten selonteossa todetaan, nykyiset tekoälysovellukset ovat heikkoa tekoälyä, koska erilaiset tekoälysovellukset eivät pysty toimimaan muissa tilanteissa. Vahva tekoäly tarkoittaisi tekoälysovelluksen laajaa muistamista ja oppimista, mikä ei ole tällä hetkellä tilanne. Tämä tarkoittaa, että

1) nykyiseen heikkoon tekoälyyn ei pidä asettaa liian suuria odotuksia ja
2) tekoälyn vaikutusten arviointi muuttuu olennaisesti kun siirrytään tarkastelemaan vahvaa, itsenäisesti oppivaa tekoälyä. Tällöin kyse on aivan eri luonteisista seurausvaikutuksista.

Tekoälyn käyttötapoihin liitetään avoimuuden käsite. Selonteossa on listattu erilaisia avoimuuden ilmenemistapoja tekoälyn käyttöön ja kehitykseen liittyen:

"Tekoälyn kehittämisessä kiinnitetään kasvavaa huomiota teknologian avoimuuteen, riskeihin ja validointiin. Avoimuuden lisäämisellä pyritään parantamaan tekoälyn hyödyntämisen turvallisuutta, hyväksyttävyyttä ja luotettavuutta. Avoimuus voi koskea tiedon käyttöä, avointa lähdekoodia, tutkimustiedon avoimuutta, turvallisuusteknologian avoimuutta, avointa innovaatiota sekä organisaatioiden tavoitteisiin ja toimintakulttuuriin liittyvää avoimuutta."

Käsitteen "avoimuus" positiivinen mielleyhtymä on hyvä asia. Samalla on kuitenkin syytä todeta, että kaikki nämä edellä mainitut käyttötapaukset eivät ole samoja ja ne eivät ole aivan yhteismitallisia. Siksi olisi tärkeää jatkossa määritellä tarkemmin, mitä tarkoitetaan.

Esimerkkinä avataan ilmaisua "tutkimustiedon avoimuus". Sillä oletettavasti tarkoitetaan vapauksia käytön suhteen, eikä pelkästään tiedon saavutettavuutta. Tällöin tulee pohtia tiedon käyttöehtoja. Peruslähtökohta on että tieto on tekijänsä omistamaa. Jos joku muu haluaa käyttää tietoa, tarvitaan lisenssi eli sopimus tai muu lupa tai oikeutus tiedon käyttämiseen. Lisenssi varmistaa tiedon käyttäjän oikeudet. Kyse ei välttämättä ole vastikkeellisesta tiedon myymisestä.

Avoimen lähdekoodin maailmassa käytetään ilmaisua copyleft ilmaisemaan sitä, mitä ehtoja lisensseissä on muutettu tavallisiin tekijänoikeuskäytäntöihin nähden ohjelmiston käyttäjän eduksi. Näin tehdään, jotta olisi selvää mitä oikeuksia käyttäjällä on. Copyleft-lisenssi (leikillisenä vastakohtana käsitteelle copyright) kuvaa poikkeuksellisen sallivat käyttöehdot. Vastaavaa mallia kannattaa harkita myös tekoälysovelluksilla käsiteltävän tiedon kohdalla ja selonteossa esitelläänkin Creative Commons -sisältölisenssit, jotka kuuluvat copyleft-lisensseihin. Nämä ovat hyvin muotoiltuja ja sertifioituja ja käytössä alan teollisuudessa. Ne ovat tehokas ja koeteltu ratkaisu tähän nimenomaiseen ongelmaan.

On kuitenkin selvää, että edellä mainittu käsitteen ”avoin” käyttötapa on hyvin erilainen kuin esimerkiksi turvallisuusteknologian avoimuudesta puhuttaessa.

Tekoälysovellusten ero sääntöperusteiseen päättelyyn ja vaikea luonnehdittavuus luonnollisella kielellä on otettava huomioon

Heikkoa tekoälyä pyörittävä kone ei ymmärrä tekevänsä päätöksiä tai suorittavansa tehtävää. Se suorittaa sitä ohjelmoidulla tavalla. Tästä syystä

tulkinta koneen suorittamasta tehtävästä on aina ihmisen tekemä. Sekä tehtävän koneelle antamisessa eli koneen ohjelmoinnissa että suorituksen tuloksen arviossa tarvitaan inhimillistä tulkintaa. Uusissa menetelmissä kuten hermoverkko-oppimisessa koneen "päättelyn" tulos ei perustukaan vanhaan tapaan, yksinkertaisten sääntöjen monimutkaisiin yhdistelmiin, vaan hermoverkko oppii saamansa datan perusteella. Hermoverkko on tietyssä tilassa opittuaan jotain. Se ei kuitenkaan osaa ilmaista luonnollisella, ihmisten puhumalla kielellä sitä, miten se on oppinut oppimansa. Hermoverkkoasiantuntijallekin se on vaikeaa siksi, että yleensä päätelmään päästään loogisista premisseistä päättelysääntöjen avulla. Hermoverkko taas ei tällaisia päättelysääntöjä käytä.

Jos tekoälyn toiminnan kuvaaminen luonnollisella kielellä on mahdotonta tai hyvin haastavaa, on vaarana, että pelkästään vihkiytyneet asiantuntijat ymmärtävät mistä on kyse. Jossain mielessä vaarana on joutua tilaan, jossa muut ymmärtävät vain sen, että "asia on näin kun kone niin sanoo". Tämä ei ole demokratian kannalta toivottavaa ja asia on otettava huomioon.

Jos käsiteltävällä asialla on laaja yhteiskunnallinen merkitys, vaikean luonnollisella kielellä kuvattavuuden voi rinnastaa luonteeltaan ja merkitykseltään toiseen, kaikkia kansalaisia koskettavaan keskusteluun: sähköiseen äänestämiseen. Sähköisen äänestämisen kansalaisyhteiskunnan ja demokratian luonteen ymmärtämisen kannalta heikoin piirre on se, että varsinainen äänestystapahtuma tapahtuu ohjelmiston sisällä ja hyvin harvalla on tietoa tai kyky ymmärtää miten äänten rekisteröinti ja laskeminen tapahtuvat. Vähintäänkin tekniikan tuntemisen vaatimukset nousevat hyvin nopeasti tavallisen äänestäjän kannalta liian korkeiksi, jotta ymmärtäminen olisi mahdollista. Paperilippuihin ja uurniin perustuvan äänestystavan ymmärtäminen on mahdollista jo päiväkoti-ikäiselle. Tällaista yksinkertaisuutta pitäisi suosia yhteiskunnassa. Tämä ei luonnollisestikaan tarkoita sitä, että kaiken toiminnan pitäisi olla selitettävissä kenelle tahansa - se olisi liiallinen vaatimus. On kuitenkin syytä pitää tällaista yksinkertaisuuden pyrkimystä hyveenä.

Tällainen yksinkertaisuuden pyrkimys tulee ulottaa myös suunnitteluun. Esimerkkinä käytetyn liikenteen palveluista annetun lain tapauksessa tietojen käsittelyn tarkoituksen korostaminen olisi tuonut esiin ajoissa senkin, koituuko halutusta tietojen käsittelystä jotain muita ei-haluttuja seurauksia. Ymmärtääksemme taksipalveluiden kilpailun vapauttamisen takia piti saada aikaan kuluttajille tapa tarkastaa kuljettajan ajo-oikeus. Toteutustavasta johtuen kuitenkin seurasi ei-toivottu sivuvaikutus: kaikkien ajo-oikeuden haltijoiden henkilötietoja, mukaan lukien henkilötunnus, oli helppo selvittää palvelun avulla. Kun tekoälyn myötä abstraktiotasoa korotetaan, mahdollisuus yhä erikoisempiin ja suurempiin virheisiin kasvaa, koska useimmiten asiantuntijat ovat keskenään sokeita täsmälleen samoille asioille.

Suomen on otettava aktiivisesti osaa alan tekniseen standardointiin

Tällä hetkellä virallinen standardointi ISO:n (International Organization for Standardization) ja IEC:n (International Electrotechnical Commission) alaisena toimiva yhteinen komitea (ISO/IEC JTC 1) kehittää tietotekniikkaan liittyviä standardeja. Tekoälyyn liittyviä standardisointeja kehittämään on perustettu oma (ISO/IEC JTC 1/SC 42) tekoälyn alakomitea, joka on tässä vaiheessa (19.12.2018) on vasta aloittelemassa toimintaansa tekoälyn standardien kehittämiseksi.

On selvää, että Suomen tulee osallistua tähän työhön aktiivisesti ja huolehtia edustuksestaan em. alakomiteassa.

Eettiset standardit ja itsesääntely eivät riitä - tarvitaan lainsäädäntöä

Tekoälysovellusten opettamisesta datalla selonteko toteaa sivulla 23:

"Tekoälyn opettamisessa tarvitaan eettisiä standardeja kansalaisen tietosuojan vahvistamisen, datan laadun sekä luotettavuuden varmistamiseen."

Effin mielestä eettiset standardit tai alan itsesäätely eivät yksistään riitä. Tarvitsemme olennaiset asiat oikeusnormeiksi ihan perinteiseen tapaan. Syynä tähän on se, että tietojen käsittely vaikuttaa voimakkasti hyvin monen, käytännössä kaikkien etuun ja siksi perussääntelyn on syytä tapahtua lainsäädännön kautta. Perussääntelyä on hyvä täydentää eettisillä standardeilla mutta esimerkiksi sen, mikä katsotaan kielletyksi on syytä olla kiellettyä laissa.

Selonteon sivulla 16 todetaan

"Tarvittaessa tietoturvallisuus taataan poikkihallinnollisin kyberturvallisuustoimenpitein."

Tämä ylätason kuvaus on liian abstrakti ilmaistakseen mitä tarkoitetaan. Koska valtion virastot ja muut viranomaistahot eivät aina suhtaudu riittävän vakavasti tietoarkkitehtuurien suunnittelemisen ja tietoturvallisen toteuttamisen osalta, tähänkin vaaditaan oikeusnormeja, jotta riittävä tietoturvan taso voidaan taata.

Esimerkkinä tällaisesta systeemisestä ongelmasta on se, että hyvin harva viranomainen on tehnyt julkisen hallinnon tietohallinnon ohjauksesta annetun lain (634/2011) 4§ velvoittaman kokonaisarkkitehtuurin suunnittelun ja kuvauksen. Kokonaisarkkitehtuurisuunnitelman ja -kuvauksen tarkoitus on ollut tarkoitus mahdollistaa hyvä suunnittelun ja toteutuksen taso ja varmistaa tietojärjestelmien yhteentoimivuus. Samalla on ollut tarkoitus saavuttaa parempi palveluiden elinkaaren hallinta ja palveluiden dokumentaatio. Näistä

vaatimuksista lipsuminen madaltaa julkisen palvelutuotannon tasoa, vaikeuttaa eheän ja mielekkään tietopolitiikan toteuttamista ja ennen kaikkea nostaa palvelutuotannon kustannuksia. On selvää, että tämä suuntaus on syytä keskeyttää ja parantaa toiminnan laatua.

Avoimet tiedostomuodot ovat yhtä tärkeitä kuin rajapinnat

Selonteko kuvailee (sivu 17) tavoitetilaa:

"Tietojärjestelmien tulisi olla mahdollisimman yleiskäyttöisiä, yhteentoimivia, tietoturvallisia ja tietoteknisesti hajautettuja. Tavoitteena on tiedon jakaminen avoimien rajapintojen kautta. Tämä ratkaisu on tietoturvallisin ja teknisen kehityksen kannalta mahdollisimman kestävä."

Jotta tietojärjestelmistä tulee mahdollisimman yhteensopivia, niissä käsiteltävän tiedon tulee olla esitettynä avoimilla tiedostomuodoilla tai -standardeilla. Jos näin ei menetellä, törmätään siihen, että lukeakseen tietoa käyttäjän täytyy hankkia ohjelmistolisenssi ohjelmistoon, joka kykenee lukemaan ja käsittelemään kyseistä tiedostomuotoa. Esimerkki tällaisesta tilanteesta on terveydenhoidossa 3-ulotteisilla kuvantamismenetelmillä otetut kuvat, erityisesti jos niissä on vielä aikaulottuvuus. Tällaisen "kuvan", oikeammin datan, esittämiseen ei ole tietääksemme avointa tiedostomuotostandardia ja siksi kansalainen ei pystyisi katsomaan omia aivokuviaan vaikka ne OmaKantaan ladattaisiinkin. Tutkimuskäytössäkin olisi pakko olla kaikki eri ohjelmistot, joilla tarkasteltavan aineiston kuvantamisia on tehty.

On yleisesti tunnettu tosiasia että liikeyritykset eivät erikoisaloilla ole motivoituneita tekemään tietojärjestelmiä tiedostomuodoiltaan avoimia. Syy on se, että samalla järjestelmästä tulisi enemmän vaihtokelpoinen muiden vastaavaan tarkoitukseen käytettävien järjestelmien kanssa. Tämä puolestaan heikentäisi niin sanottua toimittajalukkoa, joka taas on tärkeä osa eräitä alan liiketoimintamalleja. Jos katsotaan, että yhteiskunnan edun vuoksi joissain järjestelmissä on saavutettava tietojen mahdollisimman suuri vaihtokelpoisuus, yhteiskunnan on sääntelyllä velvoitettava virastot, organisaatiot ja yritykset kehittämään ja käyttämään avoimia tiedostomuotoja näissä käyttötarkoituksissa.

Kansalainen on oikeussubjekti itsenään, ei vaihdannan yksikkönä

Selonteko luonnehtii datataloustilannetta sivulla 24 seuraavasti:

"Datatalouden kehittymättömät markkinat ovat johtaneet siihen, että datatalouden suuntaa määrittää kourallinen globaaleja alustayrityksiä, joiden liiketoimintamalli perustuu datan keräämiseen ja hallinnoimiseen niiden omilla suljetuilla alustoilla."

Alustayritysten voimakas keskittyminen on syytä huomioida. Näiden muutamien yritysten määräysvalta on niin suuri, että kilpailevan tuotteen tai palvelun tuominen markkinoille on hyvin haastavaa. Moni on käytännöllisesti katsoen monopoli toimialallaan. Toinen ongelma liittyy näiden yritysten palveluiden suureen käyttäjämäärään tai paremminkin suureen osuuteen koko väestöstä. Jo tällä hetkellä on julkisia palveluita ja julkisin varoin tuotettua tietoa joka ei ole sellaisten henkilöiden saavutettavissa, jotka eivät käytä alustapalveluita. Käyttämättömyys johtuu usein tietosuojasyistä ja on tarkan harkinnan tulos. Palveluita tai sisältöjä alustoille tuottavat ihmiset tai virastotkaan eivät useimmiten hahmota sitä, että tuotettu palvelu tai sisältö on käytännössä saavuttamattomissa kyseistä alustaa käyttämättömille.

Jos tähän kehitykseen edistetään vielä tekoälyn käyttö, ollaan lähellä tilannetta, jossa kansalaisten tasa-arvo on vaarassa. Demokratian kannalta yksi olennainen kysymys on, pystyvätkö myös he, jotka eivät halua tehdä itsestään myytävää tuotetta luovuttamalla osia omasta, läheistensä ja viestintäkumppaniensa yksityisyydestä alustayritykselle, osallistumaan täysipainoisesti yhteiskunnan toimintaan. Tämä ongelma aletaan vasta tunnistaa. Sen huomioiminen tietopolitiikassa on tärkeää.

Elias Aarnio
varapuheenjohtaja
Electronic Frontier Finland – Effi ry