

Lausunto asiassa U 28/2021 vp Valtioneuvoston kirjelmä eduskunnalle komission ehdotuksesta Euroopan parlamentin ja neuvoston asetukseksi tekoälyn harmonisoiduksi sääntelyksi

Helsingin yliopiston Kuluttajatutkimuskeskus
Esittelijänä apulaisprofessori Minna Ruckenstein

1. Tavoitteet ja lähestymistapa

Sääntelyehdotuksen tavoitteena on vahvistaa eurooppalaista arvopohjaa suhteessa tekoälyjärjestelmiin ja lujittaa kansalaisten luottamusta tekoälyyn. Näiden tavoitteiden tukemana pyritään edistämään EU:n asemaa ja kilpailukykyä markkinoilla.

Sääntelyn tarkoituksena on

- 1) tunnistaa ei-toivottuja tekoälyn käyttötapoja ja tekoälyn käytänteitä ja tarvittaessa kieltää ne
- 2) identifioida korkean riskin tekoälyn käyttötapoja ja kohdistaa niihin valvontaa
- 3) edistää markkinatoimijoiden itsesääntelyä

Kielloilla pyritään varmistamaan, että markkinoilla ei toimi ei-toivottuja tekoälyjärjestelmiä. Näin kansalaiset ja tekoälyn käyttäjät voivat luottaa siihen, että EU:n alueella voi huoletta käyttää ja hankkia tekoälyjärjestelmiä.

Lisäksi sääntelytoimenpiteitä kohdistetaan tekoälyjärjestelmiin, joissa on tunnistettu yksilöihin ja yhteiskuntaan kohdistuvien haittavaikutusten riskejä.

Sääntelyehdotus on luonteeltaan reaktiivinen. Se pyrkii edistämään eurooppalaisia arvopyrkimyksiä, mutta samalla se perustuu yksinomaan riskien hallintaan. Tämä tarkoittaa sitä, että sääntelyn avulla ei pyritä aktiivisesti edistämään toivottavia tekoälykäytänteitä vaan lähinnä tunnistamaan ei-toivottuja käytänteitä ja estämään niistä syntyviä haittoja.

2. Kiellettyjen ja korkean riskin tekoälyjärjestelmien määritelmistä ja määrittelyn tavoista

Kiellettyjen ja korkeariskisten tekoälyjärjestelmien listat vaikuttavat sisältävän oikeita ja tarpeellisia asioita. Vaikutelma voi johtua siitä, että ne tuntuvat luetteloivan läpikäytyjä skandaaleita, menneitä väärinkäytöksiä ja uhkakuvaskenaarioita. Sääntely vaikuttaa siis perustuvan tuttuihin, julkisuutta saaneisiin tapauksiin.

Vaikka kielletyt ja korkean riskin käyttöalueet ovat kuluttajan luottamuksen ja turvallisuudentunteen näkökulmasta oikeansuuntaisia, määrittelyjen avoimuudesta johtuen on vaikea sanoa, mitä kaikkea sääntely pitää sisällään. Kiellettyinä listatut manipulointi, sosiaalinen pisteytys ja etätunnistusjärjestelmät herättävät joukon kysymyksiä:

- Mitä manipulointiin käytettävien tekoälyjärjestelmien kieltämisellä tavoitellaan? Sosiaalisen median palvelut pyrkivät ”alitajuisella manipuloinnilla” markkinoimaan ja tarjoamaan käyttäjille tietynlaisia sisältöjä. Facebookin kohdennustoimintoja ei ehkä ole kuitenkaan tarkoitus kieltää tällä sääntelyllä? Entä mikä lasketaan haitaksi? Onko sosiaalisen median aiheuttama ahdistus tai vaikka epäedullisten lainojen lupauksiin uskomisen tarpeeksi haitallista?
- Mitä sosiaalinen pisteytys tarkoittaa, ja milloin sen katsotaan aiheuttavan epäoikeudenmukaista kohtelua? Suuri osa julkisen sektorin etuusjärjestelmistä ja terveydenhuollon järjestelmien automaattisesta tiedonkäsittelystä perustuu jonkinlaiseen ”pisteytykseen” – halutaanko näitä olemassa olevia järjestelmiä kieltää? Jos ei, niin mitä halutaan kieltää? Kiellettyinä alueena puhutaan ainoastaan viranomaisten toimesta tai puolesta tapahtuvasta pisteytyksestä, joten yksityisen sektorin pisteytysjärjestelmät jäävät tämän kohdan ulkopuolelle. Tällä hetkellä käytössä olevat asiakkaiden pisteytys- ja luottoluokitusjärjestelmät, älyvakuutukset ym. eivät siis ole kiellon kohteena.
- Kiellettyjen asioiden listalla on biometrisen etätunnistusjärjestelmän käyttö lainvalvontatarkoituksessa julkisissa tiloissa (johon on lukuisia poikkeuksia, kuten esimerkiksi terrorismi ja turvallisuus). Yksityinen sektori voi käyttää biometristä etätunnistamista, samoin viranomaiset muuhun kuin lainvalvontaan. Lainvalvonnan poikkeuksetkin tuntuvat laajoilta, turvallisuus on monissa maissa ollut laajamittaisten valvontajärjestelmien perusteena jo pitkään. Mitä tämän kohdan on lopulta tarkoitus estää? Tällaisenaan kiello vaikuttaa ponnettomalta.

Ehdotetut kiellot ja korkean riskin järjestelmien sääntely voivat olla toimivia tapoja estää ja suitsia järjestelmiä, joiden käyttötarkoituksen tunnistamme haitalliseksi tai jollaisista meillä jo on kokemusta. Ehdotettu sääntely vaikuttaa kuitenkin tehottomammalta tunnistamaan myöhemmin ilmeneviä järjestelmien haittoja, jotka voivat olla jopa jo tunnettuja vaikutuksia haitallisempia.

Ehdotuksessa mainitaan useita kertoja, että tekoälyjärjestelmien koko elinkaari pitää huomioida. Tähän nähden on yllättävää, kuinka johdonmukaisesti sääntelyehdotus painottuu tekoälyjärjestelmien toteutusvaiheeseen. Kieltolista ei vaikuta tehokkaalta tavalta tunnistaa tekoälyjärjestelmien tahattomia pitkän aikavälin haitallisia oheisvaikutuksia, jotka havaitaan vasta sen jälkeen, kun järjestelmä on käytössä. Toisin kuin joissain muissa teknologioissa ja tuotteissa, tekoälyjärjestelmä ei ole ”valmis” kun se otetaan käyttöön, sillä sen toiminta voi muuttua ja haitalliset vaikutukset paljastuvat vasta käytön myötä. Järjestelmiä on myös tapana jatkuvasti kehittää edelleen.

3. Yleisempiä huomioita sääntelyn toteutuksesta

Ei ole uskottavaa, että tunnistaissimme kaikki tulevaisuudessa haitalliset tekoälyn käyttötavat, saati ennakoimattomat sivuvaikutukset. Olisikin tärkeää kiinnittää huomiota siihen, mihin korkean riskin tekoälyjärjestelmien listaaminen perustuu. Kuinka se on muodostettu, miten sitä päivitetään? Sekä kiellettyjen että korkean riskin tekoälyjärjestelmien

määrittelyprosessin pitäisi olla tarpeeksi avoin ja dynaaminen tunnistamaan etukäteen havaitsematta jääneitä ei-toivottuja seurauksia.

Koska sääntelyehdotus käsittelee nopeasti kehittyvää teknologian alaa, yhtä tärkeää kuin tieto tunnetuista riskialttiista käyttötarkoituksista, on menetelmä/ prosessi uusien kiellettävien kohteiden identifioimiseksi ja lisäämiseksi listalle.

Sääntelyn toimeenpano on ehdotuksessa jokseenkin epämääräinen. Ehdotuksen mukaan tekoälyjärjestelmien läpinäkyvyyden tulee toteutua niin, että ”luonnolliset ihmiset voivat valvoa niitä”. Miten tämä toteutuu käytännössä, jos järjestelmät ovat kovin monimutkaisia ja oppivat jatkuvasti uutta? Toisaalta jää myös epäselväksi, missä määrin läpinäkyvyys saavuttaa kuluttajan. Ehdotuksessa esitetään esimerkiksi vaatimus, että asiakkaalle on kerrottava, jos hänen tunteitaan tai ominaisuuksia tunnistetaan. Tämä vaikuttaa kannatettavalta tavoitteelta, mutta tieto pelkästä tunnistuksesta jää kovin vähäiseksi. Toivottavaa olisi kertoa, miten ominaisuus on tunnistettu, minkälaiseen dataan perustuen, miten tehtyä luokittelua tullaan käyttämään, ja mitä mahdollisuuksia asiakkaalla on kiistää luokittelu.

Kaiken kaikkiaan sääntelyehdotuksen pyrkimykset ovat varsin kannatettavia. On tärkeää, että tekoälyteknologioiden riskeihin ja haittoihin suhtaudutaan vakavasti. Sääntelyehdotuksen toteutus jättää kuitenkin paljon toivomisen varaa. Tulkinallisuutta ja avoimia kysymyksiä on liikaa, osin sääntelyehdotuksen esitykset vaikuttavat ajastaan jälkeenjääneeltä yritykseltä suitsia uutta teknologiaa. Sääntelyehdotuksen heikkoudet juontuvat sekä sääntelyn kohteen määrittelyn että sääntelyn toteutuksen staattisuudesta/epämääräisyydestä. Kun teknologiat kehittyvät, sääntelyn tulisi myös kehittyä paremmin teknologioihin liittyviin yhteiskunnallisiin haasteisiin vastaaviksi. Jos eurooppalaista arvopohjaa halutaan vahvistaa, sääntelyehdotusten tulisi myös herättää luottamusta ja perustua teknologiakentän syvällisempään tuntemukseen ja asioiden huolellisempaan valmisteluun.

Helsingissä 14.6.2021
Minna Ruckenstein