FCAI Finnish Center for Artificial Intelligence

Asia: U 28/2021 vp
Valtioneuvoston kirjelmä eduskunnalle komission ehdotuksesta Euroopan parlamentin ja neuvoston asetukseksi tekoälyn harmonisoiduksi sääntelyksi (Artificial Intelligence Act)

Sivistysvaliokunta 21.09.2023

Petri Myllymäki
Professor of Artificial Intelligence and Machine Learning, University of Helsinki
Vice-Director, FCAI

Aalto University

UNIVERSITY OF HELSINKI

VTT

# Tekoälysääntelyn tavoitteet

- Asetusehdotuksen olennaisena, kannatettavana tavoitteena on luoda Euroopasta ihmiskeskeisen tekoälyn kehittämiselle ja soveltamiselle suotuisa teknologisesti edistyksellinen, mutta samaan aikaan ihmisoikeuksia kunnioittava, turvallinen toimintaympäristö.
- Tähän tavoitteeseen pääsemiseksi on ensiarvoisen tärkeää, että tekoälyn sääntelyä koskeva asetus on mahdollisimman selkeä, yksiselitteinen, ristiriidaton, toimenpiteiltään oikein mitoitettu ja hyvin perusteltu.
- Koska tekoälyn alueella teknologinen kehitys on nopeaa, sääntelyn tulisi olla mahdollisimman teknologiariippumatonta
- Asetuksen nykyisessä muodossa nämä tavoitteet eivät kaikilta osiltaan toteudu, vaan tavoitetilan saavuttamiseksi on ehdotusta vielä parannettava.

# Eri sektoreiden keskeisiä huolia

- Kaupallinen sektori:
  - Mikä on asetuksen kohdealue, onko etukäteen selvää milloin tuote tai palvelu kuuluu sen piiriin?
  - Kuinka paljon työtä, aikaa ja kustannuksia asetuksen velvoitteet edellyttävät?
- Jäsenmaat/julkinen sektori:
  - Mikä on asetuksen kohdealue, mitkä kaikki palvelut kuuluvat sen piiriin?
  - Mitä seurauksia AIA:lla on jäsenmaiden omalle sääntelylle, vrt. esim. Suomen tuore laki automatisoidusta päätöksenteosta julkishallinnossa? (TFEU114, pre-emption),
- Tutkimussektori:
  - Mikä on asetuksen kohdealue, mitä seuraamuksia sillä on TKI-toiminnalle?
- On ensiarvoisen tärkeää tuoda mahdollisimman selkeästi esille se, että tarkoitus (tietenkin) on, että sääntely ei tule haittaamaan TKI-toimintaa Euroopassa, ja että korkean riskinkin käyttötarkoituksissa vaadittavat toimenpiteet tulevat olemaan järkevästi mitoitettuja ja hallittavissa
  - Tietyt tahot yrittävät luoda mielikuvia siitä, että Eurooppa on reguloimassa itsensä ulos tekoälyn kehityksestä
  - Little encouragement and no funding: why open source startup founders are leaving Europe (Sifted, July 2023)

# Asetuksen kohdealue?

- Perusongelma (”valuvika”):
  - Asetuksen tarkoitettu kohdealue on kohonneen riskin (kehittyneet) tietokonejärjestelmät, mutta asetus on rakennettu tietyn teknologian, tekoälyn, määritelmän varaan
  - Tekoälyä teknologiana ei kuitenkaan ole mahdollista määritellä, koska se on teknologiariippumaton tutkimusalue, joka käsittelee joukkoa ongelmia (puheen ymmärtäminen, tekstianalyysi, kuvantunnistus jne.), ja käytettävät menetelmät vaihtuvat aikakaudesta riippuen
- Komission ehdotuksessa määritelmä oli näennäisestä täsmällisyydestään huolimatta hyvin epäselvä, ja sisälsi listan laveita teknologiaperheitä, joita ei itsessään mitenkään määritelty
- Neuvoston ehdotuksessa määritelmää kavennettiin koskemaan vain monimutkaisia tekoälyjärjestelmiä (koneoppiminen ja sääntö/tietämyskannat), mikä selvästi jättäisi tilaa kansalliselle sääntelylle yksinkertaisilla, riskittömillä menetelmillä
- Parlamentin ehdotus on selvästi lavein:
  - **‘artificial intelligence system’ (AI system) means a machine-based system** *that is designed to operate with varying levels of autonomy and* **that can**, *for explicit or implicit objectives,* **generate outputs** *such as predictions, recommendations, or decisions that influence physical or virtual environments.*
- Määritelmä voidaan jättää laveaksi ja epämääräiseksi, kunhan asetuksessa tuodaan selvästi ilmi että sen kohdealue on kohonneen riskin järjestelmät, ja yksinkertaiset ja riskittömät tietokonejärjestelmät ovat asetuksen ulkopuolella (tällainen Suomenkin ajama tarkennus on EP:n ehdotuksesta jätetty pois)

# Perustamalleista

- Perustamallit ovat suuriin datamääriin perustuvia tekoälymalleja, joita voidaan käyttää useiden eri palveluiden tai tuotteiden perustana
  - Tällaisten mallien kehittämiskustannukset eivät ole julkista tietoa, mutta helposti miljardeista
- Tunnettuja esimerkkejä ovat mm. chatGPT, DALL-E ja Midjourney, mutta myös 50 vuotta vanhan proteiinien laskostusongelman ratkaiseva Deepmind-yhtiön julkaisema Alphafold
  - AlphaFold Is The Most Important Achievement In AI — Ever (Forbes Oct 2021)
- Erityisesti luonnollisella kielellä (ja muulla medialla) operoivat perustamallit ovat tärkeitä koska ne tarjoavat tärkeän rajapinnan kieleen ja sitä kautta kulttuuriin
  - Mikäli tällaisia malleja hallitsevat yksinomaan angloamerikkalaiset teknojätit, niiden tarjoama maailmankuva on vinoutunut, ja erityisesti pienten kielialueiden palvelut ovat toisarvoisessa asemassa
  - Oikeus osallistua tieteeseen ja kulttuuriin on yksi törkeistä yleismaailmallisista ihmisoikeuksista

# Perustamallit ja avoin ohjelmistokehitys

- Paras keino estää perustamallien monopolisoitumista ja teknojättien ylivaltaa on tukea hankkeita, joissa luodaan avoimia, vapaasti käytettävissä olevia tekoälymalleja ja ohjelmistoja, joita kehitetään yleensä suurten yhteisöllisten, vapaaehtoisvoimien varassa toimivien hankkeiden voimin
  - EU:ssa voitaisiin luoda esimerkiksi yhteinen alusta (perustamalli ja tarvittavat työkalut), jonka avulla jokainen jäsenmaa voisi kehittää hyvänlaatuiset chatGPT:n tyyliset palvelut paikallisilla kielillä
- Avoimeen ohjelmistokehitykseen osallistuvien tekijöiden joukko on niin suuri, etteivät suurimmatkaan kaupalliset teknojätit pysty samaan
  - Vuotanut Google-dokumentti: "We Have No Moat, And Neither Does OpenAI"
- Avoimet ohjelmistot ovat usein parempilaatuisia ja luotettavampia kuin suljetut/kaupalliset, koska niiden kehittämiseen ja testaamiseen osallistuu valtava joukko motivoituneita asiantuntijoita eri aloilta
  - Esimerkiksi lähes kaikki merkittävimmät webbipalvelimet ja supertietokoneet maailmassa käyttävät Linux-käyttöjärjestelmää, jota kehitetään aktiivisesti edelleen koko ajan avoimena ohjelmistona
  - Avoimet ohjelmistot tuottavat EU:lle 65-95 miljardia vuosittain
- Koska avoimet ohjelmistohankkeet ovat hajautettuja/yhteisöllisiä eivätkä yleensä ole minkään organisaation hallinnassa, niiden tuominen eurooppalaisen (tuote)sääntelyn piiriin voisi aiheuttaa merkittävää vahinkoa eurooppalaisille toimijoille, ja haitata niiden osallistumista tekoälyn kehittämiseen ja hyödyntämiseen
- Avoimella ohjelmistokehityksellä on omat sääntelymekanisminsa: esimerkiksi jos ohjelmiston dokumentointi on puutteellista, yhteisö ei osallistu sen kehittämiseen

# Perustamallit ja tekoälyasetus

- Avoin ohjelmistokehitys suljetaan EP:n ehdotuksessa selvästi sääntelyn ulkopuolelle (artikkeli 2, 5d-e), mikä on hyvä tarkennus
  - Tämä on loogisesti oikein, koska ohjelmistot (tai mallit) sinänsä eivät ole tuotteita tai palveluja, vaikka ne olisivatkin avoimesti saatavilla jonkin julkisen arkiston (esim. Github tai Hugginface) kautta
- Perustamallien suhteen tehdään kuitenkin poikkeus (5e):
  - This Regulation shall not apply to AI components provided under free and open- source licences except to the extent they are placed on the market or put into service by a provider as part of a high-risk AI system or of an AI system that falls under Title II or IV. **This exemption shall not apply to foundation models as defined in Art 3.**
- Uusi artikkeli 28b (Obligations of the provider of a foundation model): on syytä selvittää, onko tämä uusi artikkeli ja perustamallien erityiskohtelu ylipäänsä tarpeen, koska
  - Mikäli puhutaan mallista tai avoimesta lähdekoodista, ei ole syytä tuoda sitä sääntelyn piiriin
  - Mikäli taas puhutaan palvelusta joka tuodaan markkinoille esimerkiksi avoimen rajapinnan (API) kautta, kyseessä selvästi on AIA:n piiriin kuuluva palvelu, ja artikkelissa listatut velvoitteet muistuttavat hyvin paljon AIA:n korkean riskitason palvelujen velvoitteita: voisiko nämä yhdistää?
- Tarkempi analyysi ja hyviä parannusehdotuksia löytyy esimerkiksi seuraavasta artikkelista:
  - https://creativecommons.org/2023/07/26/supporting-open-source-and-open-science-in-the-eu-ai-act/