

Lausunto tekoälyn kansallisen lippulaivan FCAIn puolesta liittyen

1. komission päivitettyyn tekoälyn koordinoituun toimintasuunnitelmaan (Coordinated Plan on AI), ja
2. komission ehdotukseen Euroopan parlamentin ja neuvoston asetukseksi tekoälyn harmonisoiduksi sääntelyksi (Artificial Intelligence Act)

1. Tekoälyn koordinoitusta toimintasuunnitelmasta

Toimintasuunnitelmassa korostetaan Eurooppalaisen tekoälyn huippuosaamiskeskittymien verkoston ("AI lighthouse") olennaista merkitystä Euroopan tekoälytutkimukselle ja sen käytännön sovelluksille. On erinomaista, että komissio on tullut siihen johtopäätökseen, että tällaista tekoälyn lippulaivaa ei voi rakentaa keskityllä "Cern-mallilla", koska tekoälyn huippuosaamiskeskittymät perustuvat paitsi paikalliseen osaamiseen, myös paikallisiin yksityisen ja julkisen sektorin ekosysteemeihin, joita on mahdotonta siirtää ympäristöstä toiseen. Monikeskinen verkostomalli takaa parhaiten sen, että majakkaverkosto lisää tekoälyn impaktia laajalti koko Euroopassa, ja FCAI Suomen edustajana tässä verkostossa lisää impaktia Suomessa.

Toimintasuunnitelmassa korostetaan siis huippuosaamisen merkitystä, ja kehoitetaan jäsenmaita investoimaan tekoälyn huippuosaamiseen: *"The members states are encouraged to **set up regional and national research excellence centres around AI**, for example by using national funding instruments and RRF funds, and create a research and technology transfer structure able to attract and retain talent while at the same time aiming to become a national reference point for AI research and development. The centres would ensure regional outreach and exchange, collaborate at the European level and, together with the EU-funded networks, build the distributed European AI lighthouse; and strengthen **investment in AI research** at national level, e.g. through the RRF."*

On hyvin valitettavaa, että Suomen kansallisessa elpymisohjelmassa ei komission suosituksista poiketen investoida tekoälyn huippuosaamiseen, vaan tekoälyn osalta lähinnä sen sovelluksiin ja niitä tukeviin alustoihin. Komission toimintasuunnitelman liitteen lopussa listataan kansallisia tekoälyinvestointeja, ja niidenkin perusteella on selvää, että Suomen tulee olemaan vaikeaa säilyttää asemaansa tekoälyn edelläkävijämaana. Erityisen huolestuttavaa on, että kansallinen tekoälyn lippulaivamme FCAI, jolla on keskeinen rooli Euroopan AI-majakkaverkoston rakentamisessa, on ilman pysyvää ja tekoälyn impaktipotentialiaa vastaavaa rahoitusta, kuten toimintasuunnitelman liitteen yhteenvedosta käy ilmi.

Samoin on valitettavaa, että Suomi – jälleen toimintasuunnitelman suosituksista poiketen – ei elpymisohjelmassaan investoi digitaalisten innovaatiokeskittymien (eDIH) kansalliseen omarahoitusosuuteen (50%). Tässä olisi ollut mahdollisuus käyttää elpymisrahaa vipuna, jolla EU:lta olisi suoraan saanut saman verran lisärahoitusta Digital Europe-ohjelmasta. On myös korostettava, että eDIH-keskuksia on tarkoitus käyttää komission Digital Europe-ohjelman hankerahoituksen hankkimiseksi Suomeen, ja eDIHien kansallisen rahoitusosuuden jättäminen tyhjän päälle lisää riskiä, että eDIH-toimintaa ei voida asianmukaisesti käynnistää ja tämäkin vipuvaikutus jää hyödyntämättä.

Horisontaalisten toimenpiteiden lisäksi komissio listaa toimenpidesuunnitelmassa korkean impaktipotentialin sektoreina seuraavat: ympäristö, terveys, robotiikka, julkinen sektori, liikenne, laki/turvallisuus ja maatalous. Nämä ovat eittämättä tärkeitä yhteiskuntamme osa-alueita, mutta robotiikka ei horisontaalina teknologiana tunnu lainkaan kuuluvan tähän joukkoon, vaan tuntuu tässä yhteydessä päälleliimatulta lisäykseltä. Sen sijaan huomattakoon, että toimenpidesuunnitelmassa mainittu uusi Public-Private Partnership (PPP) sisältää alueenaan sekä tekoälyn, datan että robotiikan, ja on hyvä, että näitä teknologioita katsotaan nimenomaan horisontaalina kokonaisuutena, eikä kehitetä erillisissä siiloissa.

2. Tekoälyn sääntelystä

Asetusehdotuksen olennaisena, kannatettavana tavoitteena on luoda Euroopasta tekoälyn kehittämiselle ja soveltamiselle suotuisa teknologisesti edistysellinen, mutta samaan aikaan ihmisoikeuksia kunnioittava, turvallinen toimintaympäristö. Tähän tavoitteeseen pääsemiseksi on ensiarvoisen tärkeää, että tekoälyn sääntelyä koskeva asetukset on mahdollisimman selkeä, yksiselitteinen, ristiriidaton, toimenpiteiltään oikein mitoitettu ja hyvin perusteltu. Asetuksen nykyisessä muodossa tämä ei kaikilta osiltaan toteudu, vaan tavoitetilan saavuttamiseksi on ehdotusta vielä parannettava.

Kriittisin epäselvyys liittyy itse sääntelyn kohteeseen: asetuksessa listataan tietyt korkean riskin käyttötarkoitukset, joissa tekoälyteknologioita ei saa joko käyttää ollenkaan, tai joissa niiden käyttö edellyttää asetuksessa listattujen toimenpiteiden ottamista mukaan osaksi käyttöönottoprosessia. Tämä muotoilu luo vaikutelman, että näihin samoihin käyttötarkoituksiin on sallittua luoda digitaalisia ratkaisuja, kunhan ne eivät vaan täytä tiettyä tekoälyn määritelmää. Tämä ei voi olla asetuksen tarkoitus, vaan tarkoituksena on varmaankin sanoa, että kielletyt käyttötarkoitukset ovat kiellettyjä käytetystä teknologiasta riippumatta, ja kaikkia korkean riskin sovelluksia on tarkoituksenmukaista säädellä samalla tavalla. Taustalla lienee useita harhaluuloja ja väärinkäsityksiä:

- Toisin kuin kaiken nykyisen tekoälyhyphen keskellä saattaa luulla, korkean riskin käyttötarkoituksiin voi tehdä ratkaisuja muutenkin kuin ns. tekoälymenetelmiä käyttäen, esimerkiksi sosiaalisen pisteytyksen järjestelmiä voidaan aivan hyvin tehdä hyvinkin yksinkertaisilla menetelmillä.
- Jos asetuksessa esitetty tekoälyn määritelmä tulkitaan niin laajasti, että se kattaa ihan kaikki digitaaliset järjestelmät, tällöin ongelmaa ei ole. Tekoäly määritelläänkin asetuksessa niin laajasti ja epäselvästi, että tällainen tulkinta on jopa mahdollinen, mutta jos tämä on tarkoitus, niin miksi silloin edes luoda harhaanjohtava vaikutelma, että asetukset koskevat vain osaa digitaalisista järjestelmistä (vain tietyllä tavalla määriteltyjä tekoälyjärjestelmiä)?
- Jos ajatellaan että tietokonejärjestelmien käyttöä yleisesti säädellään tarpeeksi hyvin jo olemassaolevilla laeilla ja asetuksilla myös korkean riskin tapauksissa, herää kysymys, että miksi uutta tekoälyasetusta ylipäänsä tarvitaan, miksi samat olemassaolevat asetukset eivät koskisi yhtä lailla myös tekoälyjärjestelmiksi kutsuttuja tietokonejärjestelmiä?
- Jos kuvitellaan että perinteiset ”ei-tekoälyjärjestelmät” ovat yksinkertaisia ja luonteeltaan hyvin erilaisia kuin tekoälyjärjestelmät, niin tämä on harhaluulo: nykyiset tietokonejärjestelmät voivat olla hyvin monimutkaisia, usein vuosia tai jopa vuosikymmeniä jatkuneen kehitystyön tuloksia, jolloin myös niitä vaivaavat tekoälyjärjestelmien yhteydessä esiin nostetut ongelmat: vaikeus ymmärtää niiden toiminnan logiikkaa, ihmiskontrollin puute, epäluotettavuus jne.

On lisäksi todettava, että asetuksessa annettu **tekoälyjärjestelmien ”määritelmä” on tarkoitukseensa itse asiassa hyödytön**: liitteessä I listataan vain tekoälyn osa-alueita, joita puolestaan ei määritellä, joten kokonaisuus jää määrittelemättä, ja jää täysin avoimeksi mitkä tietokonejärjestelmät ovat asetuksen kannalta tekoälyä ja mitkä ei. Tätä asiaa on itse asiassa

mahdoton korjata koska juridiseen tarkoitukseen sopivaa tekoälyn (tai sen osa-alueiden) määritelmää ei ole olemassa: koko tieteenala muuttuu koko ajan nopeasti, ja aikoinaan tekoälyksi mielletyt menetelmät muuttuvat ajan mittaan ”normaaleiksi” digitaalisiksi järjestelmiksi.

Yksinkertainen korjausehdotus: tuodaan selkeästi esille ajatus, että kiellettyjä ja korkean riskin käyttötarkoituksia tulee säädellä teknologiariippumattomasti. Koska asetus kuitenkin tullee otsikkotasolla pysymään tekoälyasetuksena mm. PR-syistä, helppo tapa toteuttaa tämä muutos on todeta, että tässä asetuksessa **tekoälyjärjestelmä määritellään sellaisena tietokonejärjestelmänä jota voi käyttää kielletyssä tai korkean riskin käyttötarkoituksessa.** Tämäntyyppinen määritelmä korostaa käyttötarkoitusta (eikä teknologiaa), on selkeä ja helposti ymmärrettävä ja lisäksi ”tulevaisuudenkestävä”, kattaen myös kaikki tulevaisuudessa mahdollisesti kehitettävät uudet teknologiat, jotka eivät välttämättä täytä mitään nykyistä tekoälyn määritelmää.

Muita epäselvyyksiä jotka lisäävät teknologioiden kehittäjien epävarmuutta:

- Mitä käyttötarkoituksia asetus oikeastaan koskee, ja miten se vaikuttaa käytännössä? Korkean riskin alueisiin kuuluvat mm. ”essential public and private services”: miten määritellään ”essential”, ja kenen näkökulmasta? Onko esimerkiksi hakukone olennainen palvelu koska ilman sitä emme helposti löydä tietoa verkosta? Entä ovatko sosiaalisen median palvelut olennaisia henkilöille joiden toimentulo riippuu esimerkiksi heidän Youtube-näkyvyydestään?
- Millä argumenteilla listatut korkean riskin sektorit on valittu? Tämä olisi tärkeää ymmärtää, koska komissio varaa itselleen oikeuden muuttaa liitteitä (ja siis esimerkiksi riskialueiden määrittelyä) ilman uutta asetuksen käsittelyä, ja jäsenmaat eivät silloin voi enää vaikuttaa prosessiin samalla tavalla kuin asetusprosessissa.
- Asetuksessa puhutaan regulatorisesta hiekkalaatikosta, jossa tekoälysovelluksia voi kehittää hallitusti ja suojatusti, mutta tämänkin tärkeän instrumentin määrittely jätetään asetuksen ulkopuolelle myöhemmin määriteltäväksi. Kuka tällaisen ympäristön määrittelee ja toteuttaa? Miten ja milloin? Kuka sitä voi hyödyntää? Millä ehdoilla?
- Ovatko suunnitellut kontrollitoimenpiteet järkeviä ja sopivasti mitoitettuja? Asetuksessa puhutaan uusista kansallisista ja Eurooppatason toimielimistä, ja lukuisista erilaisista ohjelmistojen kehittämisprosessiin vaikuttavista vaatimuksista. Ennen asetuksen hyväksymistä olisi tärkeää varmistua siitä, että korkean riskin sovelluksiin liitettävät hallinnolliset prosessit eivät aiheuta esimerkiksi logistisia pullonkauloja hidastaen kehitystyötä, tai epärealistisia vaatimuksia liittyen kustannuksiin, työmäärään tai tekniseen seurantaan tai dokumentaatioon. Tällaisia vaikutusanalyysyjä on tekeillä, mutta niiden tuloksia ei ole vielä saatavilla - tehtävä on haastava asetustekstin monipolvisuuden ja monimutkaisuuden vuoksi.

On hyvä ja kannatettava tavoite tähdätä korkealaatuisempiin tietokonejärjestelmien tuotantoprosesseihin Euroopassa etenkin korkean riskin käyttötarkoituksissa, mutta samalla on otettava huomioon, että jos asetuksen toteutus tulee tekemään kehitystyön työläämmäksi, kalliimmaksi, hitaammaksi tai riskialttiimmaksi, tämä ei välttämättä tee Eurooppaa houkuttelevammaksi ympäristöksi. Miten luoda Eurooppaan innovaatiomyönteinen ilmapiiri jos asetuksesta saa vaikutelman että olennaisia asioita ovat riskit, uhat, sanktiot ja byrokratia, minkä vaikutelman tietyt lobbausryhmät mielellään luovat, asetuksen heikkouksia ja epäselvyyksiä korostaen (ja jopa vääristellen)?

Helsingissä 14.06.2021

Petri Myllymäki
FT, professori (tekoäly ja koneoppiminen)
Varajohtaja, Tekoälyn tutkimuskeskus FCAI