



## **Tekoälyn kehittäminen Suomessa: tekstin- ja tiedonlouhinta**

Kysymys digitaalisuudesta, tekoälystä ja datasta ovat olleet käsiteltävinä tekijänoikeusasioiden neuvottelukunnassa osana kulttuuripoliittisen selonteon valmistelua.

Teknologiateollisuus esittää tämän kysymyksen osalta seuraavat huomiot:

Datan ja tekoälyn käyttäminen ja niiden kehitys on vaikutuksiltaan ja taloudelliselta merkitykseltään erittäin merkittävä kysymys. Tekoälyä pidetään EU:ssa kriittisenä teknologiana, jolla on myös kasvava turvallisuuspoliittinen merkitys.

Suomessa on vahva traditio tekoälyn tutkimuksesta ja Suomessa on myös yksi Euroopan johtavista tekoälyratkaisuja kehittävästä eurooppalaisista yrityksistä, Silo AI, joka on kehittänyt yhteistyössä Turun yliopiston kanssa avoimeen lähdekoodiin perustuvan laajan suomenkielisen Poro-kielimallin. Suomenkielisellä datalla koulutettu malli perustuu suomalaisia edustavaan dataan, josta välittyy arvopohjamme ja kulttuurimme vivahteet. Tämä on pienelle kielialueelle harvinainen ja tärkeä resurssi.

Laajat kielimallit ja muut generatiiviset tekoälymallit koostuvat pohjimmiltaan kolmesta pääkomponentista: datasta, joka toimii mallin koulutuksen perustana, laskennasta, joka tarjoaa mahdollistaa mallien kehityksen ja koulutuksen, ja algoritmeista, jotka määrittävät, miten dataa käsitellään ja tulkitaan. Jos yksi näistä poistetaan kuviosta, ei yhtälöllä ole ratkaisua eikä synny aidosti käytettävää tekoälymallia.

Laajat kielimallit toimivat seuraavan teknologisen arvonluonnin aallon keskeisenä digitaalisena infrastruktuurina. Etenkin avoimen lähdekoodin mallit mahdollistavat erikoistuneiden mallien hienosäädön toimiala- ja käyttökohtaisiin tarpeisiin eri organisaatioiden toimesta. Tätä kautta mallit tukevat taloudellista arvonluontia ja suomenkielisten mallien tapauksessa mahdollistavat, että arvonluonti tapahtuu Suomessa.

Suomessa on hyvät edellytykset tehdä tekoälykehitystä kaikkien kielten, mutta erityisesti eurooppalaisten kielten osalta: pohjoismaisiin kieliin perustuva Viking-kielimallin kehitys on jo vauhdissa Silo AI:ssa. Kaikkia virallisia EU-kieliä kattava malli on myös valmisteilla. Suomessa on tekoälyn kehitykseen tarvittavat elementit: data, laskenta (Eurooppalainen supertietokone LUMI), algoritmit ja tarvittava osaaminen. Datan saatavuus perustuu osittain vuonna 2023 voimaan tulleeseen tekijänoikeuslain tekstin ja tiedonlouhinnan sallivaan tekijänoikeuden rajoitukseen, johon sisältyy oikeudenhaltijan oikeus pidättää tämä oikeus itsellään (opt-out).

Tulisiko tekijänoikeuslain tekstin ja tiedonlouhinnan poikkeusta muuttaa?

Nyt tekijänoikeusjärjestöt ovat aloittaneet keskustelun tekoälyn tekstin ja tiedonlouhinnan sallivan rajoituksen muuttamisesta sopimuslisenssisäännökseksi.

Sopimuslisenssillä on tarkoitus suojata teosten käyttäjää teosten massakäyttötilanteissa. Laajoja tekstiaineistoja edellyttävässä suurten kielimallien koulutuksessa ei ole kyse luonteeltaan jatkuvasta massakäytöstä vaan yleensä kerran tai harvakseltaan tehtävästä, suurteholaskentaa vaativasta toimenpiteestä. Kielto-oikeutta käyttävien oikeudenhaltijoiden kanssa aineistojen käytöstä voidaan sopia tarvittaessa suoraan.

Kehitettäessä laajaan kielimalliin perustuvia erikoistuneita kielimalleja, tapahtuu mallin kouluttaminen varsin erikoistuneella, suoraan esimerkiksi toimialan toimijoilta saatavalla aineistolla. Näissäkin tapauksissa kielto-oikeuttaan käyttävien oikeudenhaltijoiden kanssa aineistojen käytöstä voidaan sopia suoraan.

Teknologiateollisuuden näkemyksen mukaan tekoälymallien koulutusdatan osalta ei ole tarvetta ottaa käyttöön analogisen ajan massamittaisia oikeudellisia ratkaisuja, jotka tässä tapauksessa vaikuttaisivat palvelevan ainoastaan lisensoivaksi järjestöksi valittavan tekijänoikeusjärjestön asemaa.

Teknologiateollisuus pitää sopimuslisenssin soveltamista tekoälymallien koulutusdataan lyhytnäköisenä ja hyvin vaarallisena aloitteena, jolla olisi erittäin laajoja haitallisia vaikutuksia suomalaisten toimijoiden edellytyksiin kehittää strategista osaamista ja Euroopan tarvitsemia kriittisiä teknologioita ja niihin perustuvaa arvonluontia. Tekijänoikeuden kansainvälisen suojan periaatteista seuraa se, että aloitteen johdosta Suomen kieleen perustuvalla tekoälyn kehittämiseksi asetettaisiin esteitä vain Suomessa, mutta niissä maissa, joissa vastaavaa sääntelyä ei olisi, se olisi edelleen mahdollista.

Kehitettäessä tekijänoikeuslakiin perustuvaa lisensointia, Teknologiateollisuus pitää tärkeänä, että ohjenuorana on suoran sopimisen mahdollisuuksien edistäminen, johon edistyvä datan hallinta ja automatisoidut transaktiot antavat uudenlaisia mahdollisuuksia. Suomen on syytä aktiivisesti seurata kehittyviä eurooppalaisia käytäntöjä ja teknisiä ratkaisuja eikä kaavamaisesti soveltaa aivan erilaisiin käyttötapauksiin kehitettyä sopimuslisenssiä. On syytä seurata ensin minkälaiseksi EU:n tekoälyasetuksen edellyttämät käytännöt muodostuvat, joissa myös tekijänoikeussuoja on huomioitu. Eurooppalaisessa keskustelussa mm. ajatuspaja Bruegelin [raportissa](#) esitetään painavia perusteita ennemmin tekijänoikeussuojan kaventamiseen kuin laajentamiseen.

Kasettimaksujen aika on auttamatta ohi – ja se on Suomelle suuri mahdollisuus.

Helsingissä, 12 päivänä huhtikuuta 2023

Teknologiateollisuus ry

Jussi Mäkinen  
johtaja, EU-regulaatio