

HUOM! Tämä on alkuperäisen artikkelin rinnakkaistallenne. Rinnakkaistallenne saattaa erota alkuperäisestä sivutukseltaan ja painoasultaan.

Käytä viittauksessa alkuperäistä lähdettä:

Alamäki, A. 2022. Miksi tekoälylle luodaan eettisiä ohjeistuksia? Tivi, 6–7, s. 67.

PLEASE NOTE! This is an electronic self-archived version of the original article. This reprint may differ from the original in pagination and typographic detail.

Please cite the original version:

Alamäki, A. 2022. Miksi tekoälylle luodaan eettisiä ohjeistuksia? Tivi, 6–7, pp. 67.

© 2022 Alma Talent. All rights reserved.

Miksi tekoälylle luodaan eettisiä ohjeistuksia?

Tekoälyhuumassa ei pidä unohtaa, että tekoälyyn liittyy myös monia eettisiä kysymyksiä.

Useissa maissa on laadittu tekoälyn eettisiä periaatteita ja viitekehyksiä. Niitä on kuulemma jo parisensataa kappaletta. Euroopan komissio on myös aktivoitunut tässä esimerkiksi koulutuksen alueella.

Tutkimuspuolella on listattu useita eettisiä ongelmia aiheuttavia tekijöitä. Esimerkiksi tutkija **Brent Mittelstadt** kollegoineen on löytänyt kuusi ongelmallista algoritmien tuottamaa tulosta, joita avaan tässä kirjoituksessa konkreettisten esimerkkien kautta.

ENSIMMÄINEN ON SE, että tekoälyn tuottama tulos voi olla tilastollisesti oikein laskettu mutta johtopäätökset ovat vääriä. Tämä liittyy klassiseen esimerkkiin, jossa jäätelönsyönnillä ja hukumiskuolemilla on vuodesta toiseen tilastollinen yhteys, vaikkei niillä käytännössä ole yhteyttä.

Toisekseen tekoälyn tekemä tulkinta voi olla teknisesti oikein ilman minkäänlaista tilannetajua. Eräs esimerkki on kuvantunnistuksesta, jossa syväoppimisen algoritmi on tulkinnut tsunamia pakoon juoksevat ihmiset muuten vain rannalla seisoskeleiksi ihmisiksi. Tekoälyn eräs heikkous on kausaliteetin ja kontekstuaalisuuden tulkinta.

Kolmas eettinen ongelma syntyy heikosta datasta. Tätä nimitetään ”roskaa sisään roskaa ulos”-ongelmaksi. Heikkolaatuisen datan tai tilastollisesti liian pienen otoksen tuottamat tulokset voivat aiheuttaa harhaanjohtavia tulkintoja. Heikolla lähtödatalla saa luotettavuudeltaan heikkoja tuloksia. Esimerkiksi pienen nakkikioskin datalla ei voi selittää koko Suomen kaupan trendejä. Algoritmi ei voi korjata epäpätevän datan puutteita.

VINOUTUNUT KOULUTUSDATA ja sen aiheuttama syrjintä on neljäs ongelma. Tulokset ovat epäreiluja, jos niissä ei huomioida riittävästi esimerkiksi vähemmistöjen piirteitä. Maailmalta on löytynyt rasistisia esimerkkejä jopa suurten teknojättien tekoälyratkaisuista. Jos koulutusdata edustaa enemmistön omi-

naisuuksia, tekoäly voi aiheuttaa todellisessa soveltamistilanteessa esimerkiksi rasistista, uskonnollista tai sukupuolten tasa-arvoon liittyvää syrjintää.

Sosiaaliset kuplat syntyvät puolestaan siitä, että algoritmit alkavat suoltaa jollekin kohderyhmälle tätä kiinnostavaa sisältöä. Ryhmä alkaa pitää asiakokonaisuuksia totena, koska sen jäsenet eivät enää saa vastakkaisia mielipiteitä ruudulle. Algoritmit näyttävät välillisesti yhä suurempaa roolia poliittisen ilmaston syntyemisessä tai sosiaalisessa päätöksenteossa. Tätä ilmiötä nimitetään tutkimuksen alueella transformatiiviseksi efektiksi.

Kuudes haaste liittyy käytetyn datan ja algoritmien jäljitettävyyteen. Yleisesti puhutaan ”mustista laatikoista”. Eettisesti kyseenalaisia päätelmiä tai vinoumia on usein vaikea jäljittää, koska datakin voi olla jo lähtökohtaisesti manipuloitua tai vinoutunutta ilman että sitä tiedostetaan. Samoin tekoälyratkaisut ovat teknisesti monimutkaisia. Lisäksi syväoppimisessa ei edes aina tiedetä, miksi kone päätyi johonkin tulokseen. Ratkaisuna näihin perätäänkin ”selittävää” tekoälyä, joka kykenee läpivalaisemaan paremmin johtopäätökseen vaikuttaneet muuttujat ja asiat.

EETTISIÄ ASIOITA voidaan tarkastella tekoälyratkaisuja suunniteltaessa, sovellettaessa ja tuloksia hyödynnettäessä. Tekoälykeskusteluissakin on noussut esille käyttäjäkeskeinen lähestymistapa. Helpointa yksittäiselle hyödyntäjäorganisaatiolle on varmistaa käytettävän datan laatu ja algoritmien sopivuus valittuun kontekstiin.

Kaikkiin käyttötapauksiin ei välttämättä liity eettisiä näkökulmia, kun taas toisilla aloilla ne voivat olla hyvinkin keskeisiä. Esimerkkinä ovat koulutusalat, terveydenhoito, sosiaalityö ja lapsiin liittyvät palvelut.



ARI ALAMÄKI

Kirjoittaja toimii yliopettajana Haaga-Helia ammattikorkeakoulussa ja on mukana Euroopan komission datan ja tekoälyn asiantuntijaryhmässä koulutuksen alueella.

**NAKKIKIOSKIN
DATALLA EI VOI
SELITTÄÄ KOKO
SUOMEN KAUPAN
TRENDEJÄ.**