

Verkostotilaisuuden tiivistys

Paikalla parhaimmillaan 30 osallistujaa

Ohjelma:

- Miika Malin, Oulun yliopisto: Koneoppiminen ja tekoälyn opettaminen: tiedon laadun merkitys
- Antti Helin, Valtiovarainministeriö: Julkisen hallinnon automatisointi ja lainsäädäntö
- Janika Tarkoma, Tilastokeskus: Katsaus tiedon laatu- ja näkökulmiin koneoppimisessa

Tapaamisen tema: tiedon laadun merkitys automatisaatioissa, koneoppimisessa ja tekoälyn käytössä

Yhteenveto

Verkostotilaisuus koostui toivotun teeman alustuksista ja nopeasta tietoisuudesta tiedon laadun standardoinnin puolelle.

Tilaisuuden ensimmäinen puhuja, Miika Malin Oulun yliopistosta, esitteli koneoppimisen perustekniikoita: ohjattua oppimista, vahvistettua oppimista ja ohjaamatonta oppimista, joita on mahdollista käyttää myös yhdessä. Koneoppimisessa tarkastellaan useampaa tietoa-aineistoa prosessin vaiheiden edetessä. Tietoa-aineistojen laadulla on kuitenkin yhteys ja kärjistetyksi hyvälaatuisella opetusdatalla koulutettu tekoäly ei välttämättä toimi toivotusti heikompi laatuisten tietoa-aineistojen kanssa. Koneoppimisen yhteydessä metatiedot eli tieto tietosisällöstä on keskeisessä asemassa. Puuttavuus on iso haaste ja sitä pitää välttää tai korjata. Lisäksi datan pitää olla tasapainoinen siinä mielessä, että erilaiset luokat sisältävät saman verran havaintoja. Koneoppimisessa vaaditaan huomattavaa panostusta tiedon alkuperään ja sen luotettavuuteen sekä sisällön data-analyysiin, jotta yllättävistä vinoumista seuraavat virheelliset toimintamallit voidaan välttää.

Tilaisuuden toinen puhuja, Antti Helin valtiovarainministeriöstä, keskittyi aihepiiriin lainsäädäntöön. Hän nosti esiin automaattiseen ratkaisumenettelyyn ja tulevan tekoälyasetuksen soveltamiseen liittyviä keskeisiä näkökulmia. Automaattikan lisäämisen taustalla on kuitenkin tietosuoja-asetus, joka turvaa henkilöiden yksityisyyttä. Automaattista päätöksentekoa ei lähtökohtaisesti sallita yksittäiseen henkilöön kohdistuvissa päätöksissä, mutta tiedonhallintalaissa on asetettu poikkeus, jolla osassa päätöksentekoa voidaan käyttää automaattista ratkaisumenettelyä. On tärkeää huomata, että hallintolaki velvoittaa säännöt laadittavaksi ihmisen toimesta ja virkavastuulla.

Automaattisessa ratkaisumenettelyssä ilmenevien virheiden korjaamisen tulee olla ilmaista. Verkostosta nousi myös kysymys vastuista tilanteessa, jossa virhe johtuu organisaatiolta toiselle siirtyneestä virheellisestä tiedosta. Tällaisessa tilanteessa on tärkeää että tieto virheestä siirtyy myös tiedon toimittaneelle taholle ja että vastuut virheiden korjaamisesta kannattaa määritellä selkeästi.

Tilaisuuden lopuksi Janika Tarkoma Tilastokeskuksesta loi nopean katsauksen teemoihin, joita koneoppimisen tietojen laadun standardoinnissa on noussut esiin. Tiedon laatu-kehikon laatu-kriteerit on peilattu ISO 25012 -standardin laatu- ja näkökulmia vasten ja vertailusta löytyy analyysi verkkosivuilta ([linkki](#)). Uusi koneoppimisen tiedon laatu-standardi tulee julkiseksi 1.1.2024 ja painottuu vahvasti Miika Malinin puheenvuorossa korostettuihin laatu- ja näkökulmiin. Standardin pohjana on käytetty mittareita ISO 25024 -standardista, mutta lisänäkökulmia tulee sekä useamman eri prosessivaiheen tietoa-aineistotarkastelusta

että koneoppimisen lisävaatimuksista tiedon relevanssiin, monimuotoisuuteen ja tasapainoisuuteen sekä tehokkaaseen tietosisältöön ja yleisempiin teknisiin näkökulmiin, jotka mahdollistavat myös kuvaformaattien käytön.