

Tekoäly tilastotuotannossa – mahdollisuudet, rajat ja vastuu

Riitta Piela

Tilastokeskus

Pielan esityksessä Tekoäly tilastotuotannossa – mahdollisuudet, rajat ja vastuu, tarkastellaan tekoälyn roolia tilastotuotannossa mahdollisuuksien, rajojen ja Tilastokeskuksen vastuun näkökulmasta. Tekoäly ei ole uusi tekninen apuväline, vaan se kytkeytyy monin tavoin tilastotuotannon menetelmiin, prosesseihin ja laadunvarmistukseen. Generatiivinen tekoäly laajentaa tätä näkymää, koska se voi käsitellä tekstiä, tukea aineistojen analysointia, auttaa tiedon jäsentämisessä ja tuottaa luonnoksia esimerkiksi koodista, kuvauksista ja dokumentaatiosta, mutta samalla se nostaa esiin mustan laatikon ongelman: tuotosten syntyminen ja perusteet eivät ole aina täysin läpinäkyviä. Agentit voivat viedä tätä pidemmälle pilkkomalla tehtäviä osiin, käyttämällä eri työkaluja ja kuljettamalla prosessia eteenpäin asiantuntijan ohjauksessa tai täysin autonomisesti. Tutkimuksessa on esitetty erilaisia “automaattisen tilastotieteilijän” suuntaisia ratkaisuja: järjestelmiä, jotka voivat etsiä rajatussa malliperheessä aineistoon sopivia mallirakenteita, vertailla vaihtoehtoja ja tuottaa niistä ymmärrettäviä kuvauksia. Uudempi kielimallitutkimus vie ajatusta kohti mallien ehdottamista ja arviointia, mutta kyse ei vielä ole yleisestä järjestelmästä, joka itsenäisesti valitsisi oikean tilastollisen menetelmän ja vastaisi sen soveltuvuudesta.

Tekoälyn tämänhetkiset rajat ja vastuullisen käyttöönoton edellytykset on kuitenkin tunnistettava selkeästi. Toimiva koodi, uskottava ehdotus tai automaattisesti tuotettu analyysi ei vielä tarkoita, että ratkaisu olisi laadukas, perusteltu tai virallisen tilaston tuotantoon hyväksyttävä. Tekoäly voi tuottaa ehdotuksia ja välituloksia, mutta se ei kanna vastuuta siitä, mitä ne merkitsevät osana tilastotuotantoa. Erityinen haaste syntyy silloin, kun tuotantoprosessin osaksi tulee vaikeasti selitettäviä malleja tai ketjutettuja agenttiratkaisuja, joissa yksi tekoälyn tekemä ehdotus johtaa seuraavaan vaiheeseen. Näihin liittyviä riskejä voidaan hallita muun muassa selkeillä käyttörajoilla, validoinnilla, dokumentoinnilla, seurannalla ja ihmisen tekemällä hyväksynnällä kriittisissä kohdissa. Tilastokeskuksen vastuu luotettavasta tiedosta ei tällöin vähene, vaan korostuu: organisaation on kyettävä määrittelemään, mitä tekoäly saa tehdä itsenäisesti, missä tarvitaan ihmisen hyväksyntä, miten tuotoksia arvioidaan ja miten tekoälyn rooli dokumentoidaan osana tilastotuotantoa.