Markus Skrifvars
Tehohoitolääketieteen vs. professori
Anestesiologia ja tehohoito, Helsingin Yliopisto
HUS, Leikkaus- ja tehohoitokeskus
markus.skrifvars@hus.fi

PROFFAN PALSTA

Tuleeko tekoäly mullistamaan lääketieteen – olen skeptinen

Tekoäly on muodissa kaikkialla, niin tiedotusvälineissä kuin lääketiedekongressien ohjelmissa. Kun kiinalainen yritys on tuottanut uuden ja halvemman version tekoälysovelluksesta, on se pörssikursseihin vaikuttava maailmanluokan uutinen. Mantra näyttää olevan, että myös lääketiede tulee mullistumaan: tekoälyn kerrotaan pystyvän automaattidiagnostiikkaan sekä pärjäävän paremmin kuin erikoistuva lopputentissä. On ehdotettu, että tekoälyä voisi käyttää apuna Käypä hoito -suositusten laatimiseen ja näytönasteen arviointiin.

Hiljattain julkaistussa tutkimuksessa tekoälyä pyydettiin selvittämään, miksi kolme systemaattista katsausta hypotermiahoidon vaikutuksista sydänpysähdyksen jälkeiseen ennusteeseen päätyivät täysin eri loppupäätelmiin ja eri hoitosuositukseen. International Liaison Committee on Resuscitationin (ILCOR) systemaattinen katsaus ei katsonut viilennyshoitoa hyödylliseksi satunnaistetut tutkimukset huomioiden. COCHRANE päätyi täysin päinvastaiseen tulokseen: viilennyshoito lisää hyvin toipuneiden määrää, joten sitä tulisi käyttää sydänpysähdyksen jälkeen. Kolmantena mukana oli European Society of Anaesthesiology and Intensive Care Medicinen ja European Society of Emergency Medicinen hoitosuositus, joka samat tutkimukset huomioiden päätyi suosittamaan viilennyshoitoa kaikille sydänpysähdyspotilaille.

Tekoäly osaa kertoa omat kantansa, mutta ei perustella niitä.

ChatGPT- ja Gemini-kielimalleilta pyydettiin yhteenveto näistä katsauksista sekä selitys sille, miksi eri suosituksiin päädyttiin. Molemmat luettelivat katsauksien tuloksia kattavasti ja kertoivat eroavaisuuksista, joiden syytä mallit eivät kuitenkaan osanneet kertoa, vaikka aiheesta on käyty vilkasta keskustelua eri tieteellisten lehtien palstoilla. Katsausten erilaiset loppupäätelmät johtuivat mukaan valituista tutkimuksista sekä meta-analyysien tekoon käytetyistä analysointimalleista. ILCOR ei ottanut mukaan kaikkia vanhoja tutkimuksia, kuten pelkkiä kongressiabstrakteja, ja painotti analyysissa suurempia monikeskustutkimuksia. COCHRANE ei tehnyt painotusta ja otti mukaan myös 25 vuotta sitten vain yhden tutkijan toimesta tehtyjä sekä ainoastaan abstraktimuodossa julkaistuja tutkimuksia. Artikkelin kirjoittajien päätelmä oli, että ChatGPT ja Gemini tuottivat hyvän yhteenvedon, josta voisi hyvinkin olla apua hoitosuosituksen laatimisessa. Itse olen asiasta vahvasti eri mieltä, sillä tekoälymallit eivät tuoneet esiin niitä keskeisiä ongelmia tai kiistakysymyksiä, jotka olisi pitänyt huomioida päätettäessä tulisiko viilennyshoitoa suositella vai ei.

Päätin itsekin kokeilla ChatGPT:tä ja kysyin: ”Parantaako viilennyshoito ennustetta sydänpysähdyksen jälkeen?” Vastaus oli selkeä ”kyllä” perustuen vuonna 2002 suoritettuun tutkimukseen. Pyysin vielä huomioimaan uuden, paljon suuremman ja satunnaistetun tutkimuksen. ChatGPT oli kovin kohtelias, kiitti ja muutti hieman konkluusiotaan, mutta oli edelleen kallellaan, että tavoitelämmön tulisi olla alle 36 astetta. Skeptisyyteni lisäksi heräsi kysymys, miten tekoäly tähän lopputulokseen on päätynyt. Ilman alan tuntemusta minulla ei olisi mitään mahdollisuutta arvioida, onko tekoälyn antama yhteenveto oikea vai ei. Tämä ei herätä luottamusta, että tekoälyä pystyisi käyttämään tieteellisten tutkimusten näytönasteiden arvioimiseen tai hoitosuositusten laatimiseen. Jos vielä otetaan huomioon eri tekoälysovellusten harjoittama poliittinen ja muu säätely esimerkiksi sensuurin avulla, niin silloin vasta ollaankin syvissä vesissä, jonne emme toivottavasti tule lähitulevaisuudessa lääketieteen alalla ajautumaan. ■