

Koneturvallisuus ja tekoäly – suunnittelupäätöksien yksityiskohtien sparraaminen vai vastuun offloading?

2.11.2026

Tässä artikkelissa käsitellään tekoälyavusteisten työkalujen ehdotusten vaikutuksia riskienarvioinnin tarkkuuteen ja ehdotusten vaikuttamista riskienarviointia asiantuntijatyönä tekevään ihmiseen.

Työkalut ja Human-In-The-Loop

Tekoäly ja laajat kielimallit ovat levittäytyneet laajalle, kuten esimerkiksi erilaisille teollisuudenaloille, terveydenhuoltoon, liikenteeseen ja energiainfrastruktuuriin. Tämä johtuu osittain niiden kyvystä automatisoida monimutkaisia tehtäviä ja tätä kautta parantaa päätöksenteon prosesseja. Terveydenhuollossa kielimalleja käytetään kliiniseen dokumentointiin ja potilaan kanssa vuorovaikutukseen. Rahoitusalalla ne auttavat petoksen torjunnassa ja raportoinnin automatisoinnissa. Lakipuolella ne auttavat sopimusten analysoinnissa ja lakitekstien teknisessä tutkinnassa. Lisäksi niitä on käytetty apuna asiakaspalvelussa, jonka avulla on saavutettu nopeampia vastausaikoja ja tarkkuuksia, parantaen merkittävästi perinteisiä työnkulkuja (Iyenghar, 2025, s. 8).

Tekoälyavusteinen päätöksenteko on tällä hetkellä erittäin kova keskustelunaihe, ja osa ihmisistä pitäisi mahdollisena jopa tiettyjen toimintojen täyttä automatisointia (Sutinen, 2026). Kunnan henkilöiden virkavastuuseen kuuluu tietous määrittävistä rajoituksista, joihin kuuluu myös tekoälyn käyttö omissa työtehtävissä (Sutinen, 2026). Apulaisoikeusasiamiehen mukaan tekoälyn tehdessä virheitä tai hallusinoidessa vastuu on edelleen viranhaltijalla. Virkahenkilöt käyttävät julkista valtaa ja toimivat

virkavastuulla, joka sitoo heidän toimintansa lakiin, eikä tekoälyn käyttö muuta tätä vastuuasetelmaa (Sutinen, 2026).

Erittäin hyvä kysymys näiden ratkaisujen kohdalla on, kuinka paljon tekoälyavusteiset työkalut vaikuttavat päätöksentekoon, etenekin niillä aloilla missä niiden käyttöä ei ole erikseen tarkennettu tai säännelty. Osassa työkaluista on rakenteellisesti mahdollista nähdä tekoälyn tuottamaa päättelyketjua, mutta mallien nopean kehityksen takia tämä ei ole selvää. Tekoälyn sisältö on suurimmalle osalle pelkkä ”black box”, jonka toimintaa on erittäin haastavaa kuvata auki. Se, miten tekoälyavusteisen päätöksenteon järjestelmä ja järjestelmän toiminnan oikeellisuus validoidaan, tarkoittaa toistaiseksi avoinna olevaa tutkimustyötä koneturvallisuuden aihealueilla.

Kielimallien rajoittuneisuudesta, kuten hallusinoinnista, vinoumista ja yleistämisestä johtuen ihmisen kokemus tuodaan työkaluille Human-In-The-Loop-menetelmällä, jossa asiantuntija toimii osana työnkulkua (Iyenghar, 2025, s. 3). Tämä lähestymistapa on erittäin tarpeellinen turvallisuuskriittisissä sovelluksissa (mts. 3). HITL-menetelmällä pyritään vähentämään kielimallien huonoa syvällistä ymmärrystä ja epätarkkojen vastauksien tuloksia (mts. 3). Kyseinen viitekehys auttaa varmistamaan, että tekoälyn tuottamat ratkaisut ovat turvallisuuskriittisissä sovelluksissa tarkistettuja ja evaluoituja asiantuntijoiden toimesta, mikä parantaa AI-avusteisten riskianalyysien soveltuvuutta ja käytettävyyttä (mts. 3).

Muuttujat osana vinoumien tunnistamista

Tekoälyn tuottamille vinoumille voidaan rinnastaa erityisesti automaatioharha, joka tarkoittaa taipumusta yliluottaa automaatioon (Goddard ym., 2012, s. 121). Automaatioharhalle on nimetty useita synonyymejä, mukaan lukien automaation tuottama itsetyytyväisyys, liiallinen luottamus automaatioon ja vahvistusvinouma. Käyttäjien muuttujat, kuten kognitiivinen tyyli, päätöksentekoa avustavat järjestelmät ja tehtävään liittyvä aiempi kokemus, vaikuttavat automaatioharhan suuruuteen. Harhan kannalta vaikuttavia asennetekijöitä ovat luottamus ja itseluottamus. Ympäristön kautta välittävät parametrit ovat tehtävän aikainen työkuorma, tehtävän monimutkaisuus ja aikapaineet, jotka kuormittavat kognitiivisia resursseja. Harhaa lieventävinä tekijöinä mainitaan kouluttautuminen ja käyttäjän vastuun korostaminen. Työkalun itsensä kannalta harhaa lieventäviä tekijöitä ovat tekoälyn tuottaman neuvon sijainti näytöllä, päätöksentekoa avustavan järjestelmän tuloksen yhteydessä näytettävä luotettavuusarvio ja työkalu informaation lähteenä, eikä varsinaisen päätöksen avustajana (mts. 121).

Erityisesti tekoälyn kanssa havaittuun automaatioharhaan liittyvät myös tekoälyn lukutaito, ammatillisen kokemuksen taso, kognitiivinen profiili, ajan myötä kehittyvä luottamuksen dynamiikka, tarkistamisen vaatimukset ja selityksen monimutkaisuus (Romeo & Conti, 2026, s. 259). Vaikka Explainable AI (XAI) ja läpinäkyvyyden mekanismit on suunniteltu lieventämään automaatioharhaa, ylitekniset, kognitiivisesti kuormittavat tai jopa yliyksinkertaistetut selitykset voivat voimistaa perusteetonta luottamusta, erityisesti vähemmän kokemusta omaavien ammattilaisten keskuudessa, joilla ei ole kehittynyttä tekoälyn lukutaitoa. Vaikka selitykset voivat nostaa järjestelmän hyväksyttävyyden tasoa käyttäjän näkökulmasta, ovat selitykset usein riittämättömiä parantamaan päätöksen tarkkuutta tai lieventämään automaatioharhaa. Kirjoittajien mukaan käyttäjän sitoutumisen lisääminen on käytännöllisin keino vähentää automaatioharhaa

tekoälyavusteisissa työkaluissa (mts. 259).

Tehokkuuden puolesta?

Voidaan myös nähdä, että tekoäly ei ole vain suorituskykyisempi ja tarjoa useampia sovellusalueita, vaan myös täydentää ihmisen tekemää työtä ja tämä antaa suuremman potentiaalin saavuttaa täydentävää tiimin suorituskykyä (Schemmer ym., 2023, s. 411). Tämän nähdään ylittävän kummankin osapuolen yksittäisen suorituskyvyn. Kuitenkin tämäntasoinen suorituskyky voidaan saavuttaa vain täydentävien taitojen kautta. Edes siinä tapauksessa, jossa tekoälyn antama neuvo on ylivoimainen, tiimin suorituskykyä ei voi parantaa yksinkertaisesti hyväksymällä tekoälyn antaman neuvon, koska silloin ei oteta huomioon ihmisen työnkulkuun tuomaa lisäarvoa (mts. 411).

Kokemuksen mukaan tekoäly voi ensisilmäyksellä tuottaa laajempaa tietoa päätöspohjaa varten ja validoida monimutkaisia kokonaisuuksia antamalla toisen mielipiteen lopputuloksesta. Ja kaiken lisäksi erittäin nopeasti, joskus murto-osalla ajasta mitä itse tehty, perinpohjainen läpikäynti vaatii. On kuitenkin erittäin tärkeää ymmärtää tekoälyavusteisten työkalujen käytön ja julkisten kielimallipalvelujen negatiiviset vaikutukset ennen kuin lukitsee tekoälyavusteisen, koneturvallisuuteen liittyvän suunnittelupäätöksen tai mallintaa tuotantosolua sen mukaan. Virheelliseen ehdotukseen luottaminen voi myöhemmin johtaa epätoivottuun tapahtumaketjuun, jonka tarkistuksen kautta olisi voinut välttää.

Juho Pirttilahti
SEAMK

Kirjoittaja toimii TKI- painotteisena konetekniikan lehtorina Seinäjoen ammattikorkeakoululla ja opetuksen lisäksi on projektipäällikkönä Koneturva-AI ja SIX Start- hankkeissa sekä SEAMKin projektikoordinaattorina Aurora XR TechTour- hankkeessa, joka toteutetaan yhdessä Luulajan teknillisen yliopiston kanssa.

Koneturva-AI on Töysän Säästöpankkisäätiön rahoittama tutkimushanke. Hankkeen aikana kehitetään työnkulku kielimalliavusteiseen vaarojen tunnistamiseen ja riskienarviointiin koneturvallisuudessa.

Lähteet

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American medical informatics association, 19*(1), 121–127. https://doi.org/10.1136/amiajnl-2011-000089

Iyenghar, P. (2025). Clever Hans in the loop? A critical examination of ChatGPT in a human-in-the-loop framework for machinery functional safety risk analysis. *Eng, 6*(2), artikkeli 31. https://doi.org/10.3390/eng6020031

Romeo, G., & Conti, D. (2026). Exploring automation bias in human–AI collaboration: A review and implications for explainable AI. *AI & society, 41*(1), 259–278. https://doi.org/10.1007/s00146-025-02422-7

Schemmer, M., Kühl, N., Benz, C., Bartos, A., & Satzger, G. (27.–31.3.2023). Appropriate reliance on AI advice: Conceptualization and the effect of explanations. *Proceedings of the 28th international conference on intelligent user interfaces*, 410–422. https://doi.org/10.1145/3581641.3584066

Sutinen, T. (28.9.2026). Apulaisoikeusasiamies selvittää, miten kunta voi käyttää tekoälyä. *Helsingin Sanomat*. https://www.hs.fi/politiikka/art-2000012303750.html