Tekoälyverkoston taustapaperi 5:

Autonomiset järjestelmät ja niiden auditointi

Alustajat: Sasu Tarkoma / HY co-host: Sauli Eloranta / Rolls-Royce

Autonomisten toiminnallisuuksien ja järjestelmien kehittyessä ja yleistyessä tulee ensiarvoisen tärkeäksi varmistaa, että järjestelmät toimivat eettisten periaatteiden, säädösten ja sovittujen toimintamallien mukaisesti. Laadukas ja hyväksyttävä autonominen järjestelmä täyttää nämä vaatimukset sen toiminnan kaikissa tilanteissa. Nämä toimintaan liittyvät vaatimukset ovat osin kaikille järjestelmille samoja, kuten esimerkiksi eettiset ja lainsäädännöstä tulevat perusvaatimukset, ja osin ne riippuvat järjestelmän toimintaympäristöstä. Esimerkiksi tekoälyn hyödyntäminen verotuksessa, oikeusprosesseissa, lääketieteessä ja liikenteessä käsittävät toimintaympäristöön liittyviä lisävaatimuksia.

Autonomiset järjestelmät tarvitsevat siis monitasoisen laadunarviointikehikon, jossa sekä yleiset että käyttötapaukseen liittyvät vaatimukset huomioidaan. Laadunarviointikehikko mahdollistaa autonomisten järjestelmien ja tekoälyratkaisuiden laadukkuuden arvioinnin ja arviointiin pohjautuvan auditoinnin. Kaavaillun kehikon sisältämän ohjeistuksen avulla voidaan kunkin autonomisen ja tekoälyä sisältävän ratkaisun osalta selvittää järjestelmän laadukkuuteen ja riskeihin liittyvät seikat. Tekoälyn hyödyntäminen esimerkiksi yllä kuvatuissa esimerkeissä vaatii auditointia. Järjestelmien tietosuojaominaisuudet sekä toiminnan läpinäkyvyys vaativat myös auditointia.

Autonomisten järjestelmiin liittyvää riskiä voidaan arvioida laadunarviointikehikon metriikoiden avulla. On mahdollista myös automatisoida laadunarviointia, jolloin järjestelmien riskitilannetta ja auditointia voidaan tehdä jatkuvasti.

Autonomia käsittää tulevaisuudessa yhä useammin turvallisuuskriittisiä ratkaisuja esimerkiksi älyliikenteessä, terveyden- ja sairaanhoidossa sekä yhteiskunnan kriittisissä järjestelmissä. Tekoälyyn pohjautuvien ratkaisujen turvallisuus on ratkaistava osana kokonaisvaltaista järjestelmätason turvallisuusarviointia.

Standardit ja monitasoinen laadunarviointikehikko

Laadunarviointikehikko ja auditointiprosessit nojaavat vahvasti standardeihin. Onkin nähtävissä, että tekoälyjärjestelmien alueelle syntyy uusia standardeja, jotka kiinnittävät keskeisiä laatuvaatimuksia eri sovellusalueilla, esimerkiksi autonomiset autot ja laivat. Suomessa autonomisten järjestelmien kehittämistä voidaan tehostaa testiympäristöjen ja testilakien osalta, joiden avulla voidaan pilottiympäristöissä kehittää järjestelmiä ja samalla vaikuttaa standardeihin. On tärkeää, että laadunarviointikehikko ja lainsäädäntö tukevat uusien ratkaisuiden kehittämistä ja käyttöönottoa. Nähtävissä on, että tarvitaan autonomisten järjestelmien hiekkalaatikoita. jotka mahdollistavat alkuvaiheen pioneerintyötä ja kokeilua. Toiminnan laajentuessa myös vaatimukset järjestelmien toiminnalle ja laadulle kasvavat. Tarvitaan siis monitasoista laatukehikkoa.

Älyliikenteessä käytetään usein vaihtelevia määritelmiä autonomiatasoista. Yhtenäistä määritystä ei ole olemassa, ja jopa samassa sovelluksessa autonomiatasot voivat vaihdella tilanteen mukaan erittäin dynaamisesti.

Tietoturva

Autonomisten järjestelmien tietoturva on kriittinen teema ja useita hyökkäyksiä on tunnistettu esimerkiksi autonomisten autojen toiminnan estämiseen ja hämäämiseen. Autonomisten järjestelmien toiminnan takaaminen vaatii standardeja, auditointia, tietoturvauhkien riskianalyysiä ja mallinnusta sekä jatkuvaa tietoturvatestausta.

Yhteiskunnan kannalta kriittisten autonomisten järjestelmien tietoturvaan on kiinnitettävä erityistä huomiota – ja tarvittaessa toimittava vanhallakin, mutta varmaksi koetulla tekniikalla.

Tekoäly auditoinnin välineenä

Uusi mielenkiintoinen alue on tekoälyn hyödyntäminen tekoälyjärjestelmien auditoinnissa. Autonomiset järjestelmät, esimerkiksi autot, ovat monimutkaisia kokonaisuuksia, jotka koostuvat erilaisista komponenteista, esimerkiksi LIDAR tutkista, GPS-anturista ja syväoppimisalgoritmeista. Auditoiva tekoäly pystyy arvioimaan näiden komponenttien tapahtumalokia, hyödyntämään simulaatioita järjestelmän kokonaistoimintaa lokitietojen perusteella, seurata käyttäjien palautetta ja käyttötilanteita sekä lukea arvioita tuotteesta ja tuottaa näiden ja muiden tietojen perusteella malleja kokonaisuuden toiminnasta ja riskeistä. Mallien avulla voidaan tehdä auditointipäätöksiä, esimerkiksi antaa luku 1-100 kuinka hyvin kokonaisuus vastaa odotuksia ja ennustaa kokonaisuuden riskejä.

Tekoälyyn perustuvan turvallisuuskriittisen järjestelmän hyväksyttämis- ja audiotointiprosessissa tärkeänä lähtökohtana on vastaavan inhimillisen suorituksen turvallisuustason määrittely. Tämä vaatinee monessa tapauksessa uusien suoritustason mittareiden kehitystä. Vastaavasti hyväksynnän pohjaksi tarvittavan datan vaatimukset ovat epäselviä pyrittäessä kohti kokonaan uudentyyppisiä ratkaisuja. Auditointitekoälyn kehittäminen on vasta alussa ja luotettava ratkaisu vaatii merkittäviä edistysaskelia sekä perustutkimuksessa että sovelluksissa.

Kirjallisuus:

Issa, Sun, and Vasarhelyi. Editorial. Research Ideas for Artificial Intelligence in Auditing: The Formalization of Audit and Workforce Supplementation. JOURNAL OF EMERGING TECHNOLOGIES IN ACCOUNTING, Vol. 13, No. 2. 2016. Saatavilla digitaalisesti: http://aaapubs.org/doi/pdf/10.2308/jeta-10511?code=aaan-site