KOUVOLAN KAUPUNKI – TEKOÄLYPOLITIIKKA - LUONNOS

Versiohistoria

Versio	Kuvaus ja muutoksen tekijän tiedot	Muutoksen hyväksyjä	Muutos hyväksytty
0.1	Ehdotus tekoälypolitiikaksi, tekijä tietohallinto	Tietohallintojohtaja, tietosuojavastaava	
0.1	Tekoälypolitiikan luonnoksen käsittely tiedonhallinnan ohjausryhmässä	Tiedonhallinnan ohjausryhmä	Hyväksytty 9.8.2024
0.1 ja 0.11	Tekoälypolitiikan luonnoksen käsittely Kopa-joryssä 17.9.2024 ja 8.10.2024. Ulkoasun tekniset muutokset tietohallintojohtaja	Konsernipalveluiden johtoryhmä	Hyväksytty 8.10.2024
0.11	Tekoälypolitiikan ehdotuksen käsittely yhteistyöryhmässä	Yhteistyöryhmä	Hyväksytty 11.11.2024
0.11	Tekoälypolitiikka -luonnos	Kaupungin johtoryhmä	Hyväksytty 25.11.2024
0.11	Tekoälypolitiikka -luonnos	Kaupunginhallitus	

Sisällys

Johdanto

Kouvolan kaupunki määrittelee tässä politiikassa ylätasolla kaupungin linjaukset tekoälyn käytön perusteiksi. Erilaisten digitaalisten palveluiden, ohjelmien ja tietojärjestelmien käyttö on enenevässä määrin osa Kouvolan kaupungin palvelutuotantoa ja toimintaprosesseja. Politiikka toimii perustana Kouvolan kaupungin tekoälyn käyttöönottoon ja käyttöön liittyville ohjeille ja toiminnoille.

Tekoäly ja erilaiset tekoälypalvelut ovat nousseet teknologisen kehityksen suunnannäyttäjiksi sekä maailmanlaajuiseksi ilmiöksi, kun erilaisia tekoälytoteutuksia alkoi 2020-luvulla ilmestymään kaikille saatavissa olevissa kanavissa, kuten verkkosivuilla, sovelluksissa ja ohjelmistoissa. Koska kyse on teknologisen kehityksen mukanaan tuomasta edistysaskeleesta, jota on verrattu tietokoneen keksimiseen tai maailmanlaajuisen internetin julkaisuun, on selvää, että edistysaskeleen hyötyjä halutaan hyödyntää myös julkishallinnossa.

Kaikkien yksityis- tai vapaa-ajankäyttöön tarkoitettujen sovellusten tai järjestelmien käyttö viranomaisasioiden hoitamisessa tai yleisesti julkisten palveluiden tarjonnassa ei ole aina mahdollista. Kun puhumme yleisesti julkishallinnon toiminnasta sekä vielä täsmällisemmin työvälineestä, joudumme kuntaorganisaatiossa huomioimaan erilaisten järjestelmien ja ohjelmistojen tietosuojan sekä tietoturvan ennen kuin toteutusta voidaan ottaa käyttöön työvälineenä. Tämän vuoksi Kouvolan kaupungin tekoälypolitiikassa/-strategiassa on laadittu suunnitelma siitä, kuinka organisaatiossa hyödynnetään tekoälyä työvälineenä ja miten tekoälyä voidaan hyödyntää.

Tekoäly, tekoälypalvelut, tekoälyohjelmistot, generatiivinen tekoäly sekä tekoälykielet, nämä ovat vain esimerkki nimityksistä, joilla tekoälyyn yleisesti viitataan ja tämän vuoksi pelkkään tekoälyyn viittaaminen voi olla harhaanjohtavaa, sillä termistön sekä termimäärittelyn sisältö on vielä vakiintumaton. Jotta riskienhallintaa on mahdollista suorittaa, täytyy kohteen olla riittävän täsmällisesti määritelty siten, että kohteen riskejä voidaan lähestyä systemaattisesti.

Kouvolassa tekoälyllä viitataan yleisesti kaikkiin erilaisin tekoälymoottorein toteutettuihin teknisiin toteutuksiin. Tämä kattaa kaikki ne toteutukset, joissa hyödynnetään esimerkiksi autoenkooderilla (VAE) tai generatiivisella vastaverkolla (GAN) toteutetut tekoälyratkaisut. Tekstipohjaista tekoälykieltä hyödyntävinä tekoälyinä pidetään kaikkia niitä toteutuksia, jotka toimivat laajoihin kielimalleihin (LLM) perustuen (esim. ChatGPT tai llama). Lisäksi tekoälynä pidetään myös aistein havaittavia toteutuksia, kuten kuvien, äänen tai puheen muodossa tuotettuja toteutuksia (esim. Dall-e tai Midjourney -toteutukset).

Tässä dokumentissa kuvataan, miten tekoälyä hyödynnetään kaupungin tavoitteiden saavuttamiseksi ja toiminnan kehittämiseksi.

1. Tekoälyn hyödyntäminen Kouvolan kaupungilla

1.1 Tekoälytavoitteiden määrittely

Tekoälyllä on mahdollista saada hyötyjä organisaatiolle esimerkiksi päivittäisten toimintojen sujuvoittamiseksi tekstinkäsittelyssä tai asiakaspalvelussa sekä suurempien kokonaisuuksien valmistelussa ja tietomäärän hallinnassa erilaisten kehitystoimien taustatukena. Kouvolassa tekoälyn käyttö tapahtuu organisaation hyväksymien työvälineiden avulla siten, että työkäyttöön hyväksyttyjä toteutuksia käytetään ohjeistuksen mukaisesti, eikä kuluttajakäyttöön tarkoitettuja toteutuksia oteta hallitsemattomasti käyttöön.

Organisaatiossa määritetään kunkin yksikön toimesta työkäyttöön tarkoitetut työvälineet sekä dokumentoidaan käytössä olevat järjestelmät tiedonhallintamalliin, jossa ylläpidetään tietoa kaupungilla käytössä olevista ohjelmistoista, järjestelmistä ja teknisistä toteutuksista (ml. tekoälytoteutuksista). Jokaisen käyttöön otettavan tai käytössä olevan tekoälyratkaisun osalta varmistetaan, että hankkeen edistymiselle tai toteutuksen suoriutumiselle on määritelty käyttötarkoitukset ja seurantatavat siten, että työvälineen hyödyt ovat objektiivisesti todennettavissa. Hyötyjen todentamisessa voidaan käyttää esimerkiksi mitattavia päämääriä, kuten työn tehostumista, kustannuskehitystä tai palveluiden saatavuuden parantumista.

Tekoälyratkaisuita käyttöön otettaessa on aina varmistettava, että tekoälyratkaisun käytöllä noudatetaan ja tuetaan organisaation yleisiä strategisia tavoitteita.

1.2 Tekoälyn käyttäminen ja työkäytössä noudatettavat periaatteet

Tekoälyä on mahdollista käyttää useisiin erilaisiin käyttötarkoituksiin, mutta kaikenlaisten tekoälyn avulla tuotettujen dokumenttien yhteydessä ilmoitetaan, miten ja miltä osin tekoälyä on hyödynnetty. Tämä tarkoittaa esimerkiksi tekoälypohjaisten tiedotteiden tai hakukoneiden yhteydessä julkaistavaa tietoa siitä, että toteutuksessa on hyödynnetty tekoälyä. Niiltä osin, kun toiminnassa käytetään tekoälyä, kerrotaan tekoälyn käytöstä mahdollisimman avoimesti ja läpinäkyvästi.

Tekoälytoteutuksia käyttöön otettaessa on huolehdittava, että kaupungin arvoja sekä yleisiä eettisiä periaatteita noudatetaan. Kaupungin käytössä olevat työvälineet ml. tekoäly sekä kaikki muut teknologiset toteutukset ovat julkisen hallinnon ja viranomaistoiminnan työvälineitä, joten niiden käyttö perustuu kaupungin omien arvojen ja tavoitteiden lisäksi myös julkisen hallinnon yleiselle arvo- ja normiperustalle, virkamiesetiikalle ja viranomaistoimintaa säätelevälle lainsäädännölle, joissa ohjaavina periaatteina toimivat mm.:

- Avoin ja läpinäkyvä käyttö
- Oikeudenmukaisuus ja syrjimättömyys
- Tietosuoja ja yksityisyys
- Turvallisuus ja luotettavuus
- Ihmiskeskeisyys
- Eettinen suunnittelu ja käyttö
- Jatkuva seuranta ja arviointi
- Osallistuminen ja yhteistyö

1.3 Tiedonhallinta ja tekoäly

Työkäytössä hyödynnettävät tekoälytuotteet edellyttävät laadukasta tietoa toimiakseen tehokkaasti ja luotettavasti. Tämän vuoksi Kouvolassa hyödynnetään vain sellaisia tekoälysovelluksia ja -palveluita, joissa tekoälyn koulutusdatana käytetään jo julkaistua tietoa, joka on kerätty julkisista lähteistä. Kaupungin käyttämiä tekoälytuotteita ei jatkokouluteta salassa pidettävällä tai muutoin suojattavaksi määritellyillä tiedoilla, eikä tekoälytuotteita käytetä esimerkiksi automaattisessa päätöksenteossa muutoin, kuin niissä tilanteissa, joissa laki tämän erikseen mahdollistaa.

Mikäli tekoälyä hyödynnetään päätöksenteossa, kuvataan päätöksenteossa käytettävän tekoälyn koulutusta, tiedon keräämistä, säilytystä, tietojen analysointia sekä tiedonjakoa koskeva toteutus erillisellä tekoälytoteutusta käsittelevällä dokumentilla, joka julkaistaan päätöksentekoa koskevalla informatiivisella ohjeella tai verkkosivustolla esimerkiksi kaupungin verkkosivuilla. Kaupunki kuvaa dokumentissaan myös

sen, kuinka tekoälyä hallitaan sekä kuinka tekoälyllä tuotettujen päätösten laatu voidaan mitata, valvoa ja varmentaa.

Työ- tai asiakaskäytössä olevissa tekoälytuotteissa käytetyn opetusdatan alkuperän ja aitouden sekä oikeellisuuden tarkistaminen on varmistettu tai muutoin varmennettu. Varmennus voidaan toteuttaa esimerkiksi siten, että tekoälytoteutuksessa on teknisesti määritelty sen käyttävän vain siihen ennalta syötettyä tietoa tai siten, että kysymysvastaukset on teknisesti rajattu vain ennalta määriteltyihin vastausvaihtoehtoihin. Tarvittaessa vastausten varmistamiseksi voidaan myös rajata tekoälytoteutuksen toimintaa siten, ettei se hyödynnä pohjakoulutukseen käytettyä tietomassaa tai vastaa epäasiallisiin kysymyksiin.

1.4 Tekoälyn vastuullinen käyttö ja resursointi

Koska tekoälytuotteet ovat osa kaupungin työvälineistöä, joilla suoritetaan työtehtäviä, edellyttävät tekoälytuotteet sekä niiden käyttö henkilöstön koulutusta sekä jatkuvaa seurantaa. Jatkuvalla seurannalla varmistetaan, että tekoälytuotteiden käyttötarkoitukset sekä käyttötavat pysyvät hallinnassa.

Kaupungilla huolehditaan siitä, että kukin tekoälytuotteita työtehtävissään käyttävä on tutustunut työvälineeseen esimerkiksi käyttäjäkoulutuksen tai erillisten ohjeiden avulla. Kaupunki huolehtii siitä, että henkilöstölle varataan riittävästi aikaa tutustua ja perehtyä työvälineisiin sekä siitä, että heillä on jatkuvasti saatavilla tarvittavat ohjeet ja koulutukset tehtäviinsä liittyen ml. tekoälykoulutukset.

Toistaiseksi tekoälytuotteisiin liittyy riski tekijänoikeuksien alaisen, väärän tiedon tai muutoin käyttökelvottoman virheellisen tiedon antamisesta, jonka vuoksi kaupungilla suositellaan tekoälyä käytettäväksi vain tiedonhakuun, tekstin tuottamiseen tai muuhun vastaavaan omaa työtä tukevaan tehtävään, jonka lopullinen tuotos on ihmisen tarkistamaa.

Tekoälyä ei käytetä sellaisissa tehtävissä, joista tehtävää suorittavalla asiantuntijalla ei ole ymmärrystä tai osaamista tekoälyn avulla tuotetun tuotoksen oikeellisuuden tai asiasisällön arvioimiseksi. Tekoälytoteutukset ovat parhaimmillaankin vain työtä tukevia, sillä tekoäly ei kykene täysin korvaamaan ihmistä. Tämän vuoksi tekoälytuotteilla on työtehtävistä riippuva rooli työn apuna, eikä tekoälyä voida käyttää kaikissa työtehtävissä. Selvyyden vuoksi muistutetaan käyttäjiä kiinnittämään erityistä huomiota ja huolellisuutta tekoälytuotteiden käyttöön sekä annettujen ohjeistuksien noudattamiseen.

Palvelut ja yksiköt, joissa tekoälytoteutuksia hyödynnetään, vastaavat siitä, että työtehtävissä käytettävistä työvälineistä on erikseen laadittu selkeä dokumentaatio ja että tehtävistä, joissa tekoälyä ei syystä tai toisesta voida käyttää, on riittävästi ohjeistettu henkilöstöä. Tällaisia tehtäviä ovat esimerkiksi sellaiset tehtävät ja käyttötavat tai käyttötarkoitukset, joiden osalta lainsäädäntö kieltää tietojen käsittelyn tekoälyllä, tietoihin sisältyy salassa pidettäviä tietoja tai henkilötietoja, tai jos tietojen on oltava ehdottomasti oikein.

1.5 Hyväksytyt tekoälytyövälineet

Koko organisaatiota koskevaa yleispätevää ohjetta tai työvälineluetteloa kaikissa työtehtävissä käytettävistä tekoälytuotteista tai työvälineistä ei ole mahdollista tehdä, sillä kuntaorganisaation tehtäväkenttä on hyvin laaja.

Kaupungilla julkaistaan ohjeet sellaisten työvälineiden osalta, jotka kuuluvat kaikille käytössä oleviin tuotteisiin ja näiden ohjeistukset löytyvät ajantasaisina kaupungin intrasta. Kyseessä on esimerkinomainen lista niistä työvälineistä, jotka ovat avoimesti kaikkien käytettävissä. Selvyyden vuoksi todettakoon, ettei lista ole tyhjentävä, sillä erilaisia tekoälytoteutuksia tuotetaan jatkuvasti lisää ja osa näistä toteutuksista on

avoimesti internetin välityksellä hyödynnettävissä. Avointen kuluttajakäyttöön tarkoitettujen toteutusten osalta kuitenkin muistutetaan, ettei tällaisia toteutuksia ole välttämättä erikseen hyväksytty kaupungin käyttöön, joten niiden käyttäminen ei lähtökohtaisesti ole sallittua työkäytössä.

Yleisenä kaupungin tekoälylinjauksena todetaan, että työkäyttöön hyödynnettävissä tekoälytoteutuksissa voidaan käsitellä vain julkista tietoa, eikä toteutuksiin saa viedä henkilötietoja (ml. henkilötietoja sisältävät päätökset) tai muuta sellaista tietoa, joka on salassa pidettävää tai ei-julkista.

Niiltä osin, kun Kouvolan kaupungilla on käytössään erilaisia tekoälytoteutuksia eri työtehtäviin, huomioidaan tekoälytoteutuksessa seuraavat seikat:

- Pääsääntöisesti työvälineiksi hankitut ja työvälineinä käytettävät tekoälytoteutukset ovat kaupungin tiedoilla koulutettua ja tarkemmin rajattua tietoa hyödyntäviä (ei kuluttajakäyttöön tarkoitettuja kaupallisia toteutuksia).
    - Esimerkiksi KSTiedon tekoälysovellus (kaupungin Teams -ympäristössä).
- Palvelun tai yksikön työtehtävistä riippuen voidaan hyödyntää myös sellaisia tekoälytoteutuksia, joissa on julkisista verkoista kerätyillä tiedoilla tuotettuja toteutuksia (kuluttajakäyttöön tarkoitetut kaupalliset toteutukset), kun on huolehdittu toteutusta käyttöön otettaessa henkilöstön riittävästä ohjeistamisesta sekä koulutuksesta eli siitä, mitä toteutuksia ja miten kyseistä tekoälyä voidaan työtehtävissä hyödyntää.
    - Esimerkiksi verkkoselainten tekoälylaajennukset (Microsoft Bing, Copilot)
    - Selvyyden vuoksi muistutetaan, ettei kaupungin käyttäjätunnuksilla tule erikseen rekisteröityä kuluttajakäyttöön tarkoitettujen tekoälytoteutusten käyttämiseksi. Kaupunki laatii sopimukset työkäyttöön tarkoitettujen sovellusten käytöstä.

Koska tekoälyn hyödyntäminen perustuu useimmiten datan hallintaan, kiinnitetään Kouvolassa erityistä huomiota siihen, ettei koko henkilöstölle oteta käyttöön sellaisia tekoälytoteutuksia, joissa tekoälyn hyödyntämää tietoa ei voida tarkemmin rajata tai tarkistaa. Tämä kattaa mm. tekoälyn asetuksien määrittelyn sekä jatkokoulutuksen siltä osin, kuin tekoälyä on mahdollista jatkokouluttaa. Tekoälytoteutuksissa on oltava tarkasti määriteltävät asetukset, joilla voidaan erikseen huolehtia tiedon jatkokoulutuksesta sekä esimerkiksi rajoittaa tekoälylle esitettäviä kysymyksiä tai sen antamia vastauksia niin, ettei tekoäly tuota mm. rasistisia, syrjiviä tai muutoin sopimattomia vastauksia.

1.6 Tekoäly toiminnan tehostajana

Tekoälytoteutuksia voidaan hyödyntää kaupungin toiminnan tehostamiseen useilla eri työ- ja tehtäväkentillä. Kaupungissa haetaan aktiivisesti uusia keinoja tehostaa organisaation toimintaa sekä tuottaa läpinäkyvämmin toimintaamme koskevaa tietoa julkiseksi kaikkien saataville. Ohessa muutamia esimerkkejä kaupungin tavoitteista tekoälyyn liittyen:

- Läpinäkyvyyden lisääminen esimerkiksi tiedonhakumahdollisuuksia kehittämällä ja hakuominaisuuksia hyödyntämällä (esim. kaupungin päätösten vieminen tekoälytoteutuksella kuntalaisten saataville verkkoon).
- Palveluohjauksen kehittäminen sekä palveluiden löydettävyyden parantaminen (esim. tekoälypohjaiset palveluratkaisut sekä palveluista tiedottaminen)
- Yhteistyön kehittäminen erilaisten sidosryhmien ja yhteistyökumppaneiden välillä (esim. julkisten organisaatioiden, yksityisen sektorin sekä kolmannen sektorin välisen tiedonkulun kehittäminen).
- Viestinnän ja markkinoinnin kehitys tekoälytoteutuksia hyödyntämällä (esim. nopeampaa julkaisutiheyttä sekä laajempaa tiedottamista tekoälytoteutuksia hyödyntäen).

- Asiakasrajapinnan kehitys ja tiedonkulun nopeuttaminen (esim. asiakaspalautteiden käsittely ja suodatus tai luokittelu tekoälytoteutusten avulla).
- Saavutettavuus ja kielelliset toteutukset asiakaskunnan palvelulaajuuden lisäämiseksi (esim. kielenkäännösten nopeuttaminen sekä saavutettavuusvaatimusten tarkistukset tekoälytoteutuksin).

Erilaisia tekoälystä saatavia hyötyjä on täysin mahdotonta tyhjentävästi luetella, joten edellä on lueteltu vain muutamia esimerkkejä ja keskeisiä tavoitteita, jotka tukevat organisaation strategisia tavoitteita ”Yhdessä, Kestävästi, Kehittäen”.

1.7 Tekoälyn hallinta ja dokumentointi

Tekoälyn käyttö tai hyödyntäminen ei ole itsestäänselvyys tai edes välttämättömyys kuntaorganisaatioille. Tekoälyn osalta on huomattava se, ettei kuntaorganisaatio toimi trendien tai kuluttajatuotteiden ehdoilla, vaan kunta keskittyy edistämään asukkaidensa hyvinvointia ja alueensa elinvoimaa sekä järjestää asukkailleen palvelut taloudellisesti, sosiaalisesti ja ympäristöllisesti kestävällä tavalla. Tämän tehtävän onnistumiselle olennaista on tuottaa palvelut kestävällä ja kustannustehokkaalla tavalla.

Jotta Kouvola kykenee tuottamaan jatkossakin palveluita kestävällä ja kustannustehokkaalla tavalla, seurataan tekoälyn käytön vaikutuksia ja tuloksia erilaisin mittarein, kuten palveluiden käyttöasteen, tuotettujen palveluiden laajuuden sekä kustannusseurannan avulla niin, että voidaan objektiivisesti todeta tekoälyn käytön olevan hyödyllistä kunnan toiminnalle. Tekoälyn käytön vaikutuksia ja tuloksia tulee mitata ja seurata säännöllisesti sekä huomioida uusien tietojen, kokemusten tai muuttuvien olosuhteiden merkitys kullekin käytössä olevalle tekoälytoteutukselle.

Dokumentoinnin on jatkuvasti oltava ajantasaista ja riittävää, jotta tekoälyvälineiden käyttö voidaan avoimesti kuvata tehtävien ja toimintojen yhteydessä, joissa tekoälyä hyödynnetään. Dokumentointia ylläpidetään osana kaupungin tiedonhallinnan tehtäviä.

2. Tekoälyn riskienhallinta

2.1 Tekoälyyn liittyvien riskien tunnistaminen

Kouvolassa tekoälypohjaisia riskejä lähestytään riskiperusteisesti[1]. Kunkin työkäyttöön tarkoitetun tekoälytuotteen osalta laaditaan tarvittaessa riskiarvio, jossa huomioidaan ainakin seuraavat asiat:

- **Dataan liittyvät riskit:** Riskit voivat liittyä datan laatuun ja tarkkuuteen, datan yksityisyyden ja turvallisuuden suojaamiseen sekä datan eettiseen keräämiseen ja käyttöön. Epätarkka tai puutteellinen data voi johtaa virheellisiin tuloksiin ja sen väärinkäyttö yksityisyyden loukkauksiin tai muihin ongelmiin.
- **Algoritmeihin liittyvät riskit:** Riskit voivat liittyä siihen, miten tekoälyä käytetään, mutta niitä aiheuttavat myös algoritmien harhat, ylioptimointi tai ylisovitus sekä "mustan laatikon" ongelma, jossa algoritmin toimintaa on vaikea ymmärtää tai selittää.
- **Käyttöön liittyvät riskit:** Riskit liittyvät esimerkiksi tekoälyjärjestelmän väärinkäyttöön, kuten käyttämiseen haitallisiin tarkoituksiin tai käyttöön ilman riittävää ymmärrystä toiminnasta. Myös tekoälyn virheellisten vastausten hyväksyminen jatkokäyttöön on

[1] Kaupungilla hyödynnetään DVV:n riskiluokitusta tekoälyriskeistä. DVV: luokitus löytyy dokumentista ”Vinkkejä tekoälypalveluiden hyödyntämiseen, versio 1.0, VAHTI-sihteeristö 12.9.2023”. Vinkkejä+tekoälypalveluiden+hyödyntämiseen_VAHTI+hyvät+käytännöt+-tukimateriaali_1.0.pdf (dvv.fi)

käyttöön liittyvä riski. Tätä voidaan pienentää tarkastamalla vastaukset riippumattomasti, jos se on mahdollista.

- **Vastuuseen ja lainsäädäntöön liittyvät riskit:** Kuka on vastuussa, jos tekoäly tekee virheen tai aiheuttaa haittaa? Tämä on monimutkainen kysymys, johon liittyy monia oikeudellisia ja eettisiä seikkoja.
- **Kyberturvallisuusriskit:** Tekoälyjärjestelmät voivat olla haavoittuvia hyökkäyksille, kuten dataan perustuville hyökkäyksille, palvelunestohyökkäyksille tai haittaohjelmille. Näitä riskejä voi pienentää tekoälyjärjestelmän suojauksen varmistamisella ja säännöllisillä turvallisuusauditoinneilla.
- **Sosiaaliset ja eettiset riskit:** Tekoälyllä on potentiaalia muuttaa yhteiskuntaa monin tavoin, eivätkä kaikki näistä muutoksista välttämättä ole positiivisia. Tekoälyn käyttöönotossa on tärkeää ottaa huomioon sen mahdolliset vaikutukset ihmisiin ja yhteiskuntaan, mukaan lukien kysymykset syrjinnästä, oikeudenmukaisuudesta ja ihmisoikeuksista.
- **Koulutusdatan paljastumiseen liittyvät riskit:** Jos organisaatio kouluttaa tekoälyn itse omalla koulutusdatalla, joissain tapauksissa tekoälymallin kautta voidaan päästä käsiksi alkuperäiseen koulutusdataan. Jos data sisältää organisaation salassa pidettäviä tietoja, niiden paljastumisen riski on olemassa.

Edellä mainitut riskiluokat käydään systemaattisesti läpi osana tekoälytuotteiden riskien arviointia sekä kuvataan ja toteutetaan havaittuihin riskeihin sovitetut riittävät kontrollit jäännösriskin hallitsemiseksi. Jos riskiarviossa havaittuja riskejä ei voida pienentää hyväksyttävälle tasolle, ei tekoälytuotetta ole mahdollista ottaa käyttöön.

Riskiarvioiden laadinta, tietosuojan vaikutustenarviointien teko sekä muutosvaikutustenarvioinnit ovat osa organisaation jatkuvaa toimintaa ja palveluiden kehittämistä. Kehitystoimintaa tehtäessä huolehditaan siitä, että edellä mainitut tehtävät suoritetaan huolellisesti ja että kehitystoimista kertyvää dokumentointia ylläpidetään.

2.2 Tekoälytuotteissa havaittujen riskien hallinta

Kun kehitystoiminnassa on laadittu riskiarvio, jossa havaitaan erilaisia riskejä, luokitellaan riskit esimerkiksi tekoälytuotteen käyttämiseen liittyväksi riskiksi joko korkeamman tai matalamman riskitason havaintoina. Eritasoiset riskit voivat liittyä esimerkiksi seuraaviin käyttötarkoituksiin:

Korkean riskitason käyttötarkoitukset

- Julkisen vallan käyttäminen
- Henkilötietojen hyödyntäminen,
- Salassa pidettävien tietojen käyttäminen,
- Palvelun tietoturvaan ja etenkin tietosuojaan liittyvät kysymykset,
- Tekijänoikeuksiin liittyvät kysymykset,
- Tiedon oikeellisuuteen liittyvät riskit.

Matalan riskitason käyttötarkoitukset

- julkisten tietojen käsittely,
- viestintä ja tiedottaminen,
- yleisten, julkisten materiaalien tuottaminen.

On hyvin todennäköistä, että riskiarvioissa havaitaan sellaisia riskejä, joiden hyväksyminen sellaisenaan tai edes lisäkontrollien avulla voi olla organisaatiolle haasteellista. Jotta kehitystoiminnassa noudatetaan asianmukaisia toimintamalleja, muistutetaan tässä yhteydessä siitä, ettei kaupungin kehitystoiminta etene järjestelmä tai yleiset trendit edellä, vaan käyttötarveperustaisesti. Kaikkia tekoälytuotteita ei ole kaupungin tahtotilasta tai käyttötarpeesta riippumatta mahdollista käyttöönottaa, mikäli riskejä ei voida lisäkontrollien avulla hallita tai riskejä poistaa.

Kuten kaikessa kaupungin riskien hallinnassa, tulee havaitut riskit käsitellä systemaattisesti siten, että:

- Vaatimustenmukaisuus on tarkistettu (lainsäädännölliset reunaehdot selvitetty, varmistetaan että toteutus on lakien ja asetusten mukainen)
- Tietoturvaan liittyvät riskit on käsitelty ja ratkaistu (toteutuksen on oltava tietoturvallinen ja kaupungin tietosuoja- ja tietoturvapolitiikan mukainen)
- Tekniset vaatimusmäärittelyt ja teknisen toteutuksen toimivuus testattu (toteutus toimii, kuten on tarkoitus)
- Sopimukset laadittu ennen käyttöönottoa (kaupunki on sopimussuhteessa palveluntuottajaan)
- Elinkaaren hallinta on huomioitu sekä tiedon että käytettävän sovelluksen osalta (sopimuksissa huomioidaan myös elinkaaren hallinta ml. tiedon eheys ja ajantasaisuus sekä päivitystiheys ja/tai hallittu järjestelmän alasajo)

Riskienhallinnassa tulee aina erikseen huomioida edellä mainitut riskit siten, että toteutus on selvitetty etukäteen. Organisaation toiminnan kannalta voi olla haastavaa tai jopa mahdotonta jälkikäteen selvittää esimerkiksi sitä, kuinka virheellinen tieto tarvittaessa korjataan ja miten vaikkapa tekoälyn tuottama tieto tarkistetaan ennen sen julkistamista, jos tietoa ei ole selvitetty aikaisemmin.

Edellä mainitun lisäksi yksi keskeisimmistä tekoälytoteutuksiin sekä muihin julkisiin järjestelmiin, ohjelmistoihin ja palveluihin liittyvistä riskeistä liittyykin esimerkiksi sellaisen ulkopuolisen palvelutoimittajien tuottamiin palveluihin tai niiden käyttämiseen, jonka kanssa ei yleensä ole mahdollista tehdä mitään sopimusta. Esimerkiksi kuluttajakäyttöön suunnitelluissa julkisissa tuotteissa käyttö-/sopimusehdot pitää hyväksyä sellaisenaan, eikä palveluiden turvallisuutta ole mahdollista arvioida. Yksi keskeinen mahdollisuus ja keino hallita edellä mainittua riskiä on käyttää organisaation omaan tai vaikkapa julkishallinnon käyttöön erikseen rakennettua tekoälytoteutusta, esimerkiksi kotimaista, Suomessa sijaitsevaa tekoälypalvelua.

2.3 Euroopan unionin tekoälyasetus (AI act)

Tällä hetkellä erilaiset tekoälylinjaukset, strategiat tai politiikat perustuvat pitkälti voimassa olevaan lainsäädäntöön sekä tulkintaan hyvistä käytänteistä ja mahdollisista riskeistä, joita erilaisiin tekoälytoteutuksiin liittyy. Jotta erilaisten tekoälytoteutusten käyttöönotossa vältetään mahdolliset lainvastaiset tai muutoin riskialttiit toteutukset, pyritään Kouvolassa jo ennakolta huomioimaan tiedossa olevat muutokset sekä reunaehdot, joita mm. Euroopan parlamentti ja neuvosto valmistelevat. Kouvolassa huomioidaan Euroopan parlamentissa hyväksytyn tekoälyä koskevan neuvottelutuloksen keskeiset sisällöt[2] seuraavasti:

- Tekoälyjärjestelmien riskiluokittelua sovelletaan edellä kappaleessa 2.2 mainittuihin riskitasoihin:

[2] Kaupungilla hyödynnetään Euroopan parlamentissa hyväksytyn neuvottelutuloksen The Act Texts | EU Artificial Intelligence Act luonnosta, jossa hahmotetaan tekoälytoteutuksien käytön reunaehtoja.

- **Kielletyt/Ei-hyväksyttävät riskit:** Kiellettyihin tekoälytoteutuksiin kuuluvat kaikki ne toteutukset, joita pidetään tekoälyasetuksessa uhkana ihmisille. Kiellettyjä tekoälytoteutuksia ovat esimerkiksi manipulointiriskin sisältävät tekoälytoteutukset, sosiaalisen pisteytyksen mahdollistavat tekoälytoteutukset sekä biometrisen tunnistamisen toteutukset, joilla voidaan kategorisoida tai tunnistaa ihmisiä reaaliajassa.
- **Korkean riskin järjestelmät:** Korkean riskin tekoälytoteutuksia ovat toteutukset, jotka vaikuttavat kielteisesti turvallisuuteen tai yksilöiden perusoikeuksiin. Korkean riskin tekoälytoteutuksia ovat esimerkiksi EU:n tietokantaan rekisteröitävät toteutukset, jotka voivat koskea kriittisen infrastruktuurin hallintaa ja käyttöä, opetusta ja ammatillista koulutusta, työnjohtoa ja työllistymistä, maahanmuuttoa ja turvapaikka-asioita sekä lain tulkintaa tai lain soveltamista.
    - Korkean riskin vaatimuksista on mahdollisuus poiketa tietyissä määritellyissä tapauksissa.
    - Korkean riskin järjestelmiä voidaan kuitenkin käyttää, jos järjestelmä täyttää sille säädetyt vaatimukset.
- **Rajoitettujen riskien järjestelmät:** Tiettyjen järjestelmien tarjoamiselle ja käytölle asetetaan tekoälyasetuksessa avoimuusvaatimuksia, kun niiden toimintatapaan tai tuottamaan materiaaliin liittyy rajoitetumpia systeemisiä riskejä. Tällaisia ovat esimerkiksi generatiiviset tekoälytoteutukset, kuten ChatGPT. Rajoitetun riskin järjestelmissä on avoimuuden osalta varmistettava, että:
    - Tekoälyn avulla tuotetun sisällön yhteydessä ilmoitetaan sisällön olevan tekoälyllä luotua
    - Malli suunnitellaan siten, että estetään mallia luomasta laitonta sisältöä
    - Julkaistaan yhteenveto mallin koulutuksessa käytetyistä tekijänoikeudella suojatuista tiedoista
- **Matalan riskin järjestelmät:** Matalan riskin järjestelmille, kuten esimerkiksi chat- tai tekstisuodattimille, sähköpostisuodattimille, erilaisille optimointityökaluille (mm. reititys) sekä teollisuudessa käytetyille tuotantoteknisille ratkaisuille ei kohdistu tekoälyasetuksessa vaatimuksia. Matalan riskin järjestelmät arvioidaan organisaation toimintaan kohdistuvien yleisten riskien kautta.

Edellä mainitut kriteerit asettavat reunaehtoja niin tekoälytoteutuksien kehittäjille kuin myös tekoälytuotteita käyttäville tahoille. Niiltä osin, kun Kouvola hyödyntää, käyttää tai kehittää tekoälytoteutuksia käytettäväksi kaupungin toiminnassa, noudatetaan tekoälyasetuksen asettamia reunaehtoja kaikessa tekoälyn käytössä[3].

[3] Tähän lukeutuvat niin tekoälytoteutuksen käyttäjää kuin myös kehittäjää koskevat avoimuutta koskevat vaatimukset, joihin esimerkiksi luonnoksen 52 artiklassa viitataan. Part of Title IV: Transparency Obligations for Providers and Deployers of Certain AI Systems and GPAI Models (52 art.).

LÄHTEET

VAHTI – Vinkkejä tekoälypalveluiden hyödyntämiseen

Vinkkejä+tekoälypalveluiden+hyödyntämiseen_VAHTI+hyvät+käytännöt+-tukimateriaali_1.0.pdf (dvv.fi)

DVV – Turvallisen tekoälykehittämisenopas

Turvallisen+tekoälykehittämisen+opas.pdf (dvv.fi)

VM – Tekoälyn eettinen ohjeistus

Tekoälyn eettinen ohjeistus - Valtiovarainministeriö (vm.fi)

Euroopan unionin parlamentti – Tekoälyasetusluonnos (2024), tiivistelmä, hyväksytty versio

The Act Texts | EU Artificial Intelligence Act

Euroopan unionin tekoälyasetus – Tekoälyasetuksen säädöstekstin luonnos, 2024

Title I: General Provisions | EU Artificial Intelligence Act