DIGI- JA VÄESTÖTIETO-VIRASTO

Tekoälyn hallinta ja vinkit

VAHTI hyvät käytännöt -tukimateriaali

13.6.2024 versio 1.5

Sisällysluettelo

Tekoälyn hallinta ja vinkit

1. Johdanto

Tämä tukimateriaali on laadittu julkisen hallinnon organisaatioille turvallisen työskentelyn ja toiminnan edistämiseksi. VAHTI hyvät käytännöt -tukimateriaalit pohjautuvat Julkisen hallinnon digitaalisen turvallisuuden johtoryhmän (VAHTI, https://dvv.fi/vahti) asiantuntijaryhmien kokoamiin suosituksiin ja hyviin käytäntöihin riskienhallinnan, toiminnan jatkuvuuden ja varautumisen, osaamisen kehittämisen, ICT-palveluiden, tietoturvallisuuden, sekä tietosuojan alueilla. Niiden mukaan toimimalla edistämme samalla kyberturvallisuuden toteutumista.

VAHTI hyvät käytännöt -tukimateriaalit ovat ensisijaisesti suunnattu julkisen hallinnon organisaatioille, mutta ne ovat vapaasti minkä tahansa organisaation hyödynnettävissä. Toivomme, että mikäli kehitätte tai parannatte näitä materiaaleja, annatte niistä myös palautetta sisältöjen edelleen kehittämiseksi. Otamme mielellämme vastaan parannus- ja korjausehdotuksia ja julkaisemme päivitetyn version, kun niitä on kertynyt riittävästi. Voit lähettää palautetta: digiturva@dvv.fi – kirjoita otsikkoon ”VHK tekoälyn hallinta”.

Tämän dokumentin ensimmäinen versio muodostettiin hyödyntämällä muun muassa tekoälyä ja se julkaistiin 12.9.2023. Tässä versiossa 1.50 on tehty päivityksiä … xxx

Tämä 13.6.2024 julkaistu versio pohjautuu valmistelutyöryhmän ja kaikille avoimen ennakkokyselyn pohjalta päivitettyyn versioon.

Tässä materiaalissa keskitymme ennen kaikkea laajoja kielimalleja (LLM) hyödyntävien tekoälypalveluiden mahdollisuuksien ja uhkien esittelyyn, koska ne ovat nyt yleistyneet reilun 1,5 vuoden aikana merkittävässä määrin ja niiden kehittymisvauhti on ollut ennen näkemätön uusien ominaisuuksien ja mahdollisuuksien osalta.

VAHTI hyvät käytännöt tukimateriaalien tuottamisessa hyödynnämme apunamme tekoälytyökaluja, kuten on tehty tämän dokumentin aiemmissa versioissa. Kaikki sisällöt käyvät läpi tarkistus- ja kommentointiprosessin.

Julkaisun jakelu: https://dvv.fi/digiturvajulkaisut

2. Miksi tekoäly on noussut näin nopeasti ja vahvasti esille?

Tekoälystä on keskusteltu jo lähes 70 vuoden ajan.

”The term artificial intelligence was first coined by John McCarthy in 1956 when he held the first academic conference on the subject. But the journey to understand if machines can truly think began much before that. In Vannevar Bush’s seminal work As We May Think he proposed a system which amplifies people’s own knowledge and understanding. Five years later Alan Turing wrote a paper on the notion of machines being able to simulate human beings and the ability to do intelligent things, such as play Chess.”[1]

ja tämä käännettynä tekoälypalvelulla ilman ihmisen muokkausta esimerkkinä

”John McCarthy käytti termiä tekoäly vuonna 1956, kun hän piti ensimmäisen akateemisen konferenssin aiheesta. Matka sen selvittämiseksi, voivatko koneet todella ajatella, alkoi kuitenkin jo paljon ennen sitä. Vannevar Bush ehdotti uraauurtavassa teoksessaan As We May Think järjestelmää, joka vahvistaa ihmisten omaa tietämystä ja ymmärrystä. Viisi vuotta myöhemmin Alan Turing kirjoitti artikkelin, jossa hän käsitteli ajatusta siitä, että koneet voisivat simuloida ihmistä ja kykenisivät tekemään älykkäitä asioita, kuten pelaamaan shakkia.”

Tekoälypalvelut eivät suinkaan ole uusia, OpenAI julkaisi 5.3.2021 nettisivuillaan tiedotteen heidän kehittämästään GPT-kielimallista (Generative Pre-trained Transformer).

“Nine months since the launch of our first commercial product, the OpenAI API, more than 300 applications are now using GPT-3, and tens of thousands of developers around the globe are building on our platform. We currently generate an average of 4.5 billion words per day, and continue to scale production traffic.”[2]

suomennettuna

“Yhdeksän kuukautta ensimmäisen kaupallisen tuotteemme, OpenAI API:n, lanseerauksen jälkeen yli 300 sovellusta käyttää GPT-3:aa, ja kymmenet tuhannet kehittäjät ympäri maailmaa rakentavat alustamme päälle. Tuotamme tällä hetkellä keskimäärin 4,5 miljardia sanaa päivässä, ja tuotanto kasvaa edelleen.”

OpenAI julkaisi reilut 1,5 vuotta myöhemmin 30.11.2022 ChatGPT 3.5 palvelun, joka levisi vuoden 2022 lopussa ennätysmäisen nopeasti ja käynnisti uuden aikakauden tekoälyn hyödyntämisessä.

Palvelusta julkaistiin 14.3.2023 versio 4.0. Version julkaisemisen jälkeen palveluun on lisätty mm. plugin-toiminnallisuus, joka mahdollistaa erilaisten rajapintojen hyödyntämisen.

[1] The History of Artificial Intelligence (washington.edu)
[2] GPT-3 powers the next generation of apps (openai.com)

13.5.2024 OpenAI julkaisi version 4o-version, johon on luvattu kesäkuun 2024 aikana merkittäviä uusia ominaisuuksia, kuten kehittyneen vuorovaikutteisen keskustelun, tuen 50 kielelle, multimodaalisen käyttöliitymän, joka mahdollistaa myös vuorovaikutteisesti videokuvan sisällön aukipurkamisen sekä kehittyneen tulkkaustoiminnon. Valtaosa näistä ominaisuuksista on luvattu myös ohjelman ilmaisversion käyttäjille. [3]

Edellä oleva esimerkki on poimittu OpenaAI:n kehittämästä tekoälypalvelusta, mutta samalla pitää muistaa, että vastaavanlainen kehitys on käynnissä kaikkien merkittävien tekoälypalveluita kehittävien yritysten osalta, esimerkiksi

Google Gemini https://gemini.google.com
Meta https://www.meta.ai/ - ei käytössä vielä Suomessa
Microsoft CoPilot https://copilot.microsoft.com
-ilmaisia ja mm. pilvipalveluihin kytkeytyviä versioita

Näiden ohella sadat, tuhannet yritykset ovat lisänneet omiin palveluihin tekoälypalveluiksi luokiteltavia toimintoja, esimerkiksi chatbot-tyyppisiä avustimia tai liittymiä laajoihin kielimalleihin.

Viimeisen vuoden aikana erityisesti kuvien tuottamisessa on saavutettu merkittävää laadullista parantumista sekä samaa loikkaa on seuraavaksi odotettavissa tekoälyllä tuotettujen videoiden osalta.

Samoin odotettavissa on, että luonnollisen kielen käyttäminen tekoälypalveluiden käyttöliittymänä tulee kasvamaan, jolla saattaa olla merkittäviä vaikutuksia esimerkiksi meille kaikille tuotettavien palveluiden tuottamisessa.

2.1 Tekoäly on yksi merkittävimmistä teknologisista kehitysaskelista

IBM julkaisi henkilökohtaisen tietokoneen elokuussa 1981. Tietokoneet toisiinsa yhdistävät tietoliikenneverkot ja etenkin internet-verkko ja nettiselaaminen 1990-luvulla ovat olleet merkittäviä ATK:n ja ICT:n kehitysaskeleita. 2000-luvulla mobiiliteknologioan ja älylaitteiden yleistyminen ovat muokanneet tapaamme käyttää digitaalisia palveluita ajasta ja paikasta riippumatta.

Tekoälyn merkitystä on mahdotonta vielä tässä vaiheessa kokonaan ymmärtää ja arvioida. Yksi Osmo W. Wiion laeista koskee tulevaisuuden ennustamista: ”Lähitulevaisuus yliarvioidaan ja kaukainen tulevaisuus aliarvioidaan.” Meillä saattaa olla käsillä hetki, jossa viimeiset pari vuosikymmentä yliarvioituamme tekoälyn tuomia mahdollisuuksia, olemmekin pian aliarvioimassa sen lähitulevaisuuden merkitystä, vaikka kuinka pyrkisimme ymmärtämään sitä.

Keskeisiä tekoälyn mahdollistajia ovat ICT-teknologian suorituskyvyn kehittyminen (laskentateho, tietoliikenneyhteyksien nopeus, tallennusjärjestelmien kapasiteetti)

[3] Introducing GPT-4o and more tools to ChatGPT free users | OpenAI

sekä teknologisten laitteiden hintatason merkittävä lasku. Tekoälypalvelut eivät olisi mahdollisia ilman ns. **Mooren lain** toteutumista[4]

Mooren laki, joka on nimetty Intelin, keskeisen mm. mikroprosessoreita valmistavan yrityksen yhden perustajan Gordon Mooren mukaan, väittää, että transistoreiden lukumäärä kaksinkertaistuu noin joka toinen vuosi, joka samalla karkeasti yleistettynä myös tuplaa tietokoneiden suorituskyvyn. Se on ollut vuosikymmenien ajan melko tarkka ennuste, mutta asiantuntijapiireissä on yhä enemmän keskustelua siitä, milloin ja miten tämä trendi tulee päätökseensä. Nykyisin monet arviot olettavat, että Mooren lain toteutuminen voisi päättyä 2020-luvun loppuun mennessä, mutta tästä on olemassa paljon epävarmuutta.

Joitakin teknologioita, kuten **kvanttitietokoneita** tai **uusia puolijohdemateriaaleja**, tutkitaan mahdollisina keinoina jatkaa suorituskyvyn kasvua sen jälkeen. On tärkeää huomata, että vaikka transistorien tiheys ei enää kasvaisikaan, tietotekniikan edistys voi jatkua muilla tavoin. Esimerkiksi **ohjelmistoalgoritmit**, **tekoäly** ja **erikoistuneet piirit** (kuten grafiikkaprosessointiyksiköt tai tekoälypiirit)[5] parantavat suorituskykyä tai tehostaa laskentaa, vaikka laitteiston suorituskyky ei enää kaksinkertaistuisikaan joka toinen vuosi.

Erinomainen esimerkki tästä Yhdysvaltalainen Nvidia-yritys [6], josta on tullut yksi maailman nopeimmin kasvavia on teknologiayrityksiä käytännössä viimeisen parin vuoden aikana. He ovat pystyneet luomaan erityisesti tekoälylaskentaa vauhdittavia, alun perin peli- ja vaativaan graafiseen kuvankäsittelyyn tarkoitetuista näytönohjainprosessoreista erityisen tehokkaita tekoälyprosessoreita.

Lisäksi tekoälyn kehittymistä on edistänyt datan, tiedon määrän merkittävä kasvu sekä tekoälyn käyttämien algoritmien kehittyminen.

Eräs tiedonmäärän kasvuun liittyvä haaste liittyy tekoälyllä luotavan tiedon määrän kasvuun. Jos kerta tekoäly ei varsinaisesti luo mitään uutta, onko vaarana se, että meidän käytettävissä oleva tieto ”tasapäistyy” tai että meistä ihmisistä tulee tyhmempiä? EI välttämättä, koska tekoäly tarjoaa meille uudenlaisia näkökulmia ja sitä kautta pystymme toimimaan hyödyntäen keskeistä meidän ihmisten voimavaraa, luovuuttamme, myös uusien innovaatioiden, ideoiden ja teorioiden tuottamiseksi.

Vaikka vuoden 2022 loppupuolelta käynnistynyt tekoälybuumi ei jatkuisi tulevina vuosina samalla intensiteetillä, se on kuitenkin käynnistänyt laajan muutoksen. Mikäli uusien ja olemassa olevien palveluiden kehittäminen hidastuisi, se ei tule pysäyttämään käynnistynyttä muutosta, mutta nyt tätä kirjoitettaessa kesällä sellaista ei näytä olevan ihan heti näköpiirissä.

[4] What Is Moore's Law and Is It Still Relevant in 2023? (makeuseof.com)
[5] Tekoälylaskenta väittää päihittävänsä Mooren: Nvidia and the AI Hardware Race - Is Moore's Law Dead (factored.ai)
[6] Nvidia - Wikipedia

3. Vinkkejä henkilöstölle tekoälypalveluiden testaamiseen ja hyödyntämiseen

Alla olevat ohjeet on tarkoitettu huomioitavaksi avointen, julkisesti saatavilla olevien, geneeristen, erityisesti laajoihin kielimalleihin (LLM) pohjautuvien tekoälypalveluiden hyödyntämisessä. Mikäli organisaatio ottaa käyttöönsä erikseen hankkimansa palvelun, sen tulee ohjeistaa palvelun käytöstä. Myös silloin on suositeltavaa tarkistaa tässä esille nostetut asiat.

3.1 Viisi keskeistä asiaa tekoälyn hyödyntämisessä

Ennen kuin ryntäät hyödyntämään tekoälyä tai viimeistään nyt alkuinnostuksen laannuttua, huomioi seuraavat neljä asiaa:

1. Ymmärrä, mitä tekoäly on ja kuinka se toimii
2. Ylläpidä omaa osaamistasi palveluiden käyttäjänä
3. Tunnista, mitä tietoja palveluun voi syöttää
4. Tarkista palveluiden tuottama tieto ennen sen julkaisua tai hyödyntämistä
5. Varaudu verkkorikollisten ja muiden vihamielisten toimijoiden väärinkäytöksiin

3.1.1 Ymmärrä, mitä tekoäly on ja kuinka se toimii

Tekoälyn toiminnan ymmärtäminen helpottaa arvioimaan niitä tekijöitä, sekä riskejä että mahdollisuuksia, joita tekoälypalveluiden käyttämiseen liittyy. Tässä tukimateriaalissa keskitymme vahvasti laajojen kielimallien (LLM) avulla toteutettuihin tekoälypalveluihin, koska ne ovat nyt nostaneet tekoälyn hyödyntämisen kokonaan uudelle tasolle.

OpenAI ChatGPT-4 määrittelee itse ”Tekoäly (Artificial Intelligence, AI) on tietojenkäsittelytieteen ala, joka keskittyy älykkäiden koneiden, ohjelmistojen ja järjestelmien kehittämiseen. Tekoälyn tavoitteena on luoda järjestelmiä, jotka kykenevät suorittamaan tehtäviä, jotka vaativat normaalisti ihmisen älykkyyttä. Näitä tehtäviä voivat olla esimerkiksi kuvan ja puheen tunnistaminen, oppiminen, suunnittelu, ongelmanratkaisu ja päätöksenteko.”

Koska tekoälyjen arkkitehtuuri ja toimintamalli voivat käytetystä palvelusta ja mallista riippuen poiketa merkittävästi, organisaation tulee ymmärtää ja tunnistaa kyseinen toimintamalli perusteellisesti. Yhden toimintamallin ja palvelun ymmärrys ei takaa sitä, että samat opit pätevät johonkin toiseen palveluun.

Tekoälypalvelu vaatii toimiakseen seuraavia asioita:

Data
Tekoäly ja koneoppiminen perustuvat suurelta osin dataan. Se voi olla esimerkiksi historiallista, reaaliaikaista, julkisesti saatavilla olevaa tai yrityksen sisäistä dataa. Se voi olla myös jossain nettipalvelussa olevaa, käyttöoikeuksien takana olevaa dataa, esimerkiksi sosiaalisen median palvelussa. Datatyyppi voi olla strukturoitua (esim. taulukkomuodossa) tai strukturoimatonta (esim. kuvat, teksti).

Algoritmit (toimintalogiikka)
Algoritmit ovat koneoppimisen ja tekoälyn sydän. Ne oppivat datasta malleja ja suhteita sekä tekevät sen perusteella ennusteita tai päätöksiä. Eri tarkoituksiin on olemassa erilaisia algoritmeja. Tavallisen ICT-palvelun yhteydessä algoritmi toimii aina samalla tavalla, esimerkiksi taulukkolaskennassa 1+1 = 2 ja 4*5=20. Sen sijaan tekoälypalveluiden osalta tulos pitäisi olla oikein, mutta on mahdollista, että kerran sata kertaa kysyttäessä, palvelu antaa tuloksen 3 tai 25. Tekoälypalveluiden hyödynnettävyyttä rajoittaakin merkittävästi se, että niiden tuottama tietoa pitää aina tarkistaa.

Kannattaa myös muistaa, että vastauksiin vaikuttaa aikaisemmin käyty keskustelu. Jatkossa on odotettavissa, että kehittyneemmissä palveluissa on käytettävissä muisti, joka saattaa jatkossa muistaa kaiken sen kanssa käydyn keskustelun.

Koneoppimisen mallit
Algoritmit käyttävät dataa koneoppimisen mallien kouluttamiseen. Mallit ovat niitä, jotka uuden syötteen saadessaan tekevät varsinaiset ennusteet tai päätökset.

Opetusdata
Tekoälyn opetusdata koostuu niistä tiedoista, joita käytetään tekoälymallin kouluttamiseen. Tämä data toimii opetusmateriaalina, jonka avulla malli oppii tunnistamaan kuvioita, tekemään ennusteita tai suorittamaan muita tehtäviä. Opetusdatan laatu (oikeellisuus) ja määrä ovat kriittisiä tekijöitä mallin toiminnan kannalta. Opetusdata on vain yksi osa tekoälyn koulutusprosessia, mutta hyvin kriittinen.

Laskentaresurssit
Tekoälypalvelut tarvitsevat usein merkittäviä laskentaresursseja, erityisesti suurten datamäärien käsittelyyn ja koneoppimismallien kouluttamiseen. Tämä voi tapahtua joko paikallisesti tai pilvipohjaisissa palveluissa.

Käyttöliittymä
Tekoälypalvelun käyttäjät tarvitsevat tapoja kommunikoida palvelun kanssa. Tämä voi tapahtua esimerkiksi web-käyttöliittymän, mobiilisovelluksen tai API-rajapinnan kautta, jatkossa yhä useammin myös puheella, kuvien tai videon avulla.

Arkkitehtuuri ja palveluiden tuottamiseksi tarvittava ICT-infrastruktuuri
Tekoälypalvelu tarvitsee taustalleen hyvin suunnitellun arkkitehtuurin ja ICT-infrastruktuurin, joka mahdollistaa tehokkaan datan käsittelyn, laskennan ja palvelun käyttämisen. Ne voivat sisältää tietokantoja, palvelimia, verkkoyhteyksiä ja muita teknologioita.

Tietoturva ja tietosuoja, yksityisyyden suoja
Tekoälypalveluiden on otettava huomioon tietoturva ja tietosuoja, erityisesti silloin kun ne käsittelevät henkilötietoja. Esimerkiksi tietojen salaus, käyttöoikeudet ja mahdollinen tietosuojan alaisen materiaalin pseudonymisointi tai anonymisointi mahdollistavat tietoturvan ja tietosuojan toteutumista. Erityisen tärkeää on tietää, päätyykö tekoälypalvelussa käsitellyt tiedot osaksi palvelun opetusdataa vai ei. Tämä vaikuttaa merkittävästi siihen, voidaanko palvelussa käsitellä esimerkiksi salassa pidettäviä tai henkilötietoja. Organisaatio voi hankkia käyttöön sellaisen tekoälypalvelun, joka mahdollistaa myös edellä mainittujen tietojen käsittelyn turvallisesti.

Tekoälyetiikka
Tekoälyetiikka tarkoittaa periaatteita ja ohjeita, jotka ohjaavat tekoälyn kehittämistä ja käyttöä siten, että ne ovat eettisesti hyväksyttäviä ja edistävät yhteiskunnan hyvinvointia. Tekoälyetiikka pyrkii varmistamaan, että tekoälyteknologiat kehitetään ja käytetään tavalla, joka on reilu, turvallinen, läpinäkyvä ja vastuullinen.

- *Reiluus ja tasa-arvo*
    - Varmistetaan, että tekoälyjärjestelmät eivät syrji käyttäjiä esimerkiksi rodun, sukupuolen, iän tai muiden henkilökohtaisten ominaisuuksien perusteella.
- *Läpinäkyvyys*
    - Tekoälyn toiminnan tulee olla selitettävissä ja ymmärrettävissä, jotta käyttäjät ja sidosryhmät voivat ymmärtää, miten ja miksi tekoäly tekee tiettyjä päätöksiä.
- *Vastuullisuus*
    - Tekoälyn kehittäjien ja käyttäjien tulee olla vastuullisia tekoälyn käytöstä ja sen vaikutuksista. Tämä tarkoittaa myös sitä, että käytössä on mekanismeja vastuun ottamiseksi, jos tekoäly aiheuttaa haittaa.
- *Turvallisuus ja luotettavuus*
    - Tekoälyjärjestelmien tulee olla turvallisia käyttää, ja niiden tulee toimia luotettavasti erilaisissa olosuhteissa. Tämä sisältää myös riskienhallinnan ja -arvioinnin.
- *Yksityisyys ja tietosuoja*
    - Tekoälyn tulee kunnioittaa yksilöiden yksityisyyttä ja varmistaa, että henkilökohtaiset tiedot käsitellään asianmukaisesti ja turvallisesti.
- *Hyvän tekeminen ja haitan välttäminen*
    - Tekoälyjärjestelmien tulisi edistää yhteiskunnan hyvää ja minimoida mahdolliset haittavaikutukset. Tämä tarkoittaa tekoälyn kehittämistä ja käyttämistä tavoilla, jotka hyödyttävät mahdollisimman monia ihmisiä.

Ota huomioon myös nämä seikat tekoälypalveluita hyödyntäessäsi:

Tekoälyn koulutus

Kielimallit on "koulutettu" syöttämällä sille suuria määriä tekstiä, jotka sisältävät monenlaisia kirjoja, artikkeleita ja verkkosivuja. Koulutuksen aikana tekoäly oppii tilastollisesti tunnistamaan, miten sanat ja lauseet liittyvät toisiinsa ja miten ne muodostavat

merkityksellisiä viestejä. Tämän ohella koulutukseen saattaa kuulua myös ihmistyönä tehtyä lisäkoulutusta, jossa karsitaan pois vastauksia, joita tekoälyn omistava taho ei syystä tai toisesta halua mallin tuottavan. Käyttäjän on siten tärkeä ymmärtää, että tekoäly ei tuota vastauksia suoraan aineiston perusteella, vaan tulokseen vaikuttavat muun muassa palvelua tarjoavan yrityksen arvovalinnat, liiketoimintalogiikka ja sovelluttuva lainsäädäntö. Tulos ei siis ole välttämättä neutraali. Sama kannattaa huomioida myös hakukoneita käytettäessä, koska hakutuloksiin voidaan vaikuttaa monilla eri keinoilla.

Tekstin tai muun vastauksen tuottaminen

Kun tekoäly saa syötteen (prompt), kuten kysymyksen tai lauseen, se yrittää oppimansa perusteella valita seuraavan sanan, joka sopii tilastollisesti parhaiten tekstiyhteyteen. Tekoäly toistaa tätä prosessia uudestaan ja uudestaan, kunnes se on tuottanut koko vastauksen tai tekstikappaleen.

Tekoälyllä ei ole ymmärrystä tai tietoisuutta

Vaikka tekoälypalvelu voi tuottaa tekstiä, joka näyttää ymmärrettävältä tai tietoiselta, on tärkeää ymmärtää, että tekoälyllä ei todellisuudessa ole ymmärrystä tai tietoisuutta. Sen tuottama tieto voi näyttää loogisesti oikealta ja uskottavalta, mutta sen sisältö voi olla täyttä puppua. Tämän takia on erittäin tärkeää, että tekoälyn tuottaman tiedon tarkistaa ja varmistaa riittävän asiantuntemuksen omaava asiantuntija. Se ei "tiedä" tai "ymmärrä" informaatiota samalla tavalla kuin ihminen. Se vain luo malleja sen pohjalta, mitä se on nähnyt aiemmin ja joka on matemaattisesti lähimpänä oikeaa vastausta.

Tekoälypalvelu toimii yksinkertaistettuna samankaltaisesti kuin älylaitteissa käytetty ns. ennustava tekstinsyöttö, joka pyrkii ennustamaan kirjoittamasi sanan syöttämiesi merkkien perustella. Se vain ennustaa yhden sanan sijaan merkittävästi laajempia kokonaisuuksia sille koulutetun aineiston perusteella.

Rajoitukset

Tekoälypalveluiden käyttöön liittyy rajoituksia. Se ei aina tuota täysin tarkkaa tai luotettavaa tietoa, ja vastaukset saattavat vaihdella, vaikka kysymys olisi sama. Se ei myöskään välttämättä muista aiemmin tapahtuneita keskusteluita, eikä sillä ole kykyä ymmärtää tai arvioida ihmisen tunteita.

3.1.2 Ylläpidä omaa osaamistasi palveluiden käyttäjänä

Joudumme nykyaikana harvoin opettelemaan täysin uudenlaisia digipalveluita tai työkaluja, sillä osalla meistä Windows, internet, sähköposti ja toimisto-ohjelmat ovat olleet käytössämme jopa vuosikymmeniä. Niiden uudet versiot tuovat harvoin mukanaan merkittäviä muutoksia.

Sen sijaan tekoälypalvelut tulevat kehittymään ennen näkemättömällä vauhdilla, jonka takia jokaisen kannattaa osallistua aiheeseen liittyviin koulutuksiin sekä aktiivisesti tutustua palveluiden tarjoamiin uusiin tai kehittyviin mahdollisuuksiin. Jos jokin,

niin tekoälyn hyödyntäminen kannustaa hyödyntämään digirohkeutta; ei uhkarohkeasti, vaan uhat tunnistaen ja riskit halliten!

3.1.3 Tunnista, mitä tietoja palveluun voi syöttää

Älä koskaan syötä tekoälypalveluihin arkaluonteista, henkilötietoa tai salassa pidettävää tietoa, ellei palvelua ole organisaatiossasi sellaisen käyttöön luokiteltu. Suositeltavaa on välttää myös muun organisaatioon liittyvän, tunnistettavan tiedon syöttämistä, jos ei ole tiedossa, miten palvelu sitä käsittelee. Tällöin on nimittäin vaarana, että tällainen tieto voi päätyä esimerkiksi osaksi palvelun opetusdataa ja paljastaa ulkopuolisille tietoja organisaation toiminnasta.

Tunnista se uhka, että jos palvelun opetusdataan päättyy jotain organisaatiokohtaista, esimerkiksi sen nimellä varustettua tai muuten tunnistettavaa dataa, joku toinen voi saada tämän tiedon käyttöönsä kysymällä suoraan ”Millainen strategia yrityksellä xyz on?”

Varmista, että osaat tunnistaa ja erottaa palveluissa hyödynnettävät julkiset tiedot kielletyistä tiedoista. Voit hyödyntää tietojen turvallisessa käsittelyssä 4T-mallia (Tunnista Tiedot, Tilat ja Työkalut) – linkki materiaaliin. Tässä mallissa tekoälypalvelut voidaan nähdä työkaluna muiden digipalveluiden joukossa.

3.1.4 Tarkista tekoälypalveluiden tuottama tieto ennen sen julkaisua ja hyödyntämistä

Muista tarkistaa tekoälypalveluiden tuottamat tulokset. Vaikka tekoäly voi automatisoida monia prosesseja, sen tulokset pitää aina tarkistaa ennen niiden julkaisemista tai muuta hyödyntämistä.

Palveluiden virheet saattavat heijastua myös niiden tuottamien tietojen eettisinä ongelmina. Varmista, että palvelut tai niiden tuottamat tiedot eivät edistä syrjintää tai epäoikeudenmukaisuutta.

Huomaa, että tekoälyä ei voi ilman erillistä hyväksyntäprosessia käyttää julkisessa päätöksenteossa, koska automaattinen päätöksenteko vaatii esimerkiksi käytettyjen mallien ja algoritmien julkisuutta. Tämä ei monesti ole mahdollista tekoälyalgoritmien kohdalla, koska malli muokkautuu koulutusmateriaalin mukaan eikä ole aina selkeästi saatavilla. Tähän liittyvää lainsäädäntöä ollaan parhaillaan kehittämässä.

Ota talteen myös käyttämäsi kehotteet (prompt), joiden avulla olet tuottanut materiaalia (tekstiä, kuvia, videoita). Sinulla saattaa myöhemmin tulla tarve tarkistaa, kuinka olet saanut tuotettua kyseisen tiedon tekoälypalvelun avulla.

Koska tekoälypalveluiden toiminta on usein suljettua, niiden taustalla olevia algoritmeja ei yleensä ole mahdollista tarkistaa. Tällaisten "mustan laatikon" tuottamien tietojen oikeellisuudesta tulee olla erityisen huolellinen.

3.1.4.1 Mikä voi aiheuttaa hallusinointia?

Tekoälypalvelut saattavat tuottaa ns. ”hallusinaatioita”, joissa palvelun tuottama tieto vaikuttaa periaatteessa oikealta, mutta käytännössä se voi olla täysin virheellistä. Tuloksessa saatetaan viitata sellaisiin lähteisiin, jotka vaikuttavat oikeilta, mutta joita ei löydykään tai ne sisältävät virheitä.

Kontekstin puute

Tekoälymallit eivät aina ymmärrä täysin laajempaa kontekstia tai tiettyjen kysymysten taustaa. Tämä voi johtaa siihen, että ne tarjoavat tietoja, jotka kuulostavat oikeilta, mutta eivät ole tarkkoja tai paikkansapitäviä.

Rajoitettu tiedon saatavuus

Tekoälymallit koulutetaan valtavista tietomassoista, mutta ne eivät voi tietää kaikkea. Jos mallilla ei ole tarpeeksi relevanttia tietoa tietystä aiheesta, se saattaa täyttää puutteet keksimällä tietoa ikään kuin ”omasta päästään”.

Tilastollinen päättely

Tekoälymallit perustuvat tilastolliseen päättelyyn. Ne ennustavat seuraavan sanan tai lauseen perustuen aiempiin malleihin, mutta ne eivät tiedä, onko ennustus oikeasti totta vai ei.

Moniselitteisyys tai monitulkintaisuus

Luonnollinen kieli on moniselitteistä, ja monet sanat ja lauseet voivat tarkoittaa eri asioita eri konteksteissa. Tämä voi johtaa siihen, että malli tuottaa vastauksia, jotka ovat teknisesti oikeita, mutta aivan väärässä kontekstissa.

Koulutusdata

Tekoälymallit koulutetaan valtavilla tietoaineistoilla, jotka sisältävät sekä oikeaa että virheellistä tai epätarkkaa tietoa. Malli ei aina pysty erottamaan näitä toisistaan, mikä voi johtaa virheellisten tietojen tuottamiseen.

Generatiiviset mallit

GPT-mallit ovat generatiivisia, mikä tarkoittaa, että ne luovat vastauksia lennossa ennakoiden, mikä voisi olla järkevää vastauksena. Tämä prosessi ei kuitenkaan takaa, että luotu tieto on totta.

Puutteelliset ohjeet

Jos käyttäjän kysymys on epäselvä tai monitulkintainen, malli voi "arvata" vastauksen perustuen aiempiin samankaltaisiin kysymyksiin. Tämä voi johtaa virheelliseen

tietoon.

3.1.5 Varaudu verkkorikollisten ja muiden vihamielisten toimijoiden hyökkäyksiin ja väärinkäytöksiin

Tekoälyn hyödyntäminen palveluissa ja laitteissa tarjoaa meille uudenlaisia mahdollisuuksia kehittää nykyistä vuorovaikutteisempia ja käyttäjäystävällisempiä palveluita ja laitteita. Sen tarjoamien ihmiskuntaa hyödyntävien mahdollisuuksien rinnalla tulee kuitenkin ymmärtää, että sen tarjoamat mahdollisuudet ovat yhtä hyvin ja osin helpommin hyödynnettävissä verkkorikollisten ja muiden väärinkäyttäjien toimesta.

Kaikki ne positiiviset mahdollisuudet, joita voimme tekoälypalveluiden avulla toteuttaa, voidaan siten kääntää meitä vastaan. Tässä tulee huomioida, että verkkorikollisia eivät koske mitkään lait tai eettiset velvoitteet ja sen takia on odotettavissa, että lähivuosien aikana koemme yhä useampia tekoälyä laajasti hyödyntäviä huijauskampanjoita, muita hyökkäyksiä ja loukkauksia. Ohessa muutamia esimerkkejä niistä osa-alueista, jotka tarjoavat tekoälypalveluiden avulla toteutettuna nykyistä kehittyneempiä tai uudenlaisia väärinkäytösmahdollisuuksia:

3.1.5.1 Rajaton mahdollisuus tuottaa ja analysoida tekstiä eri kielillä laadukkaasti

Digihuijauskampanjoiden toteuttaminen siten, että

- viestin sisältö on mahdollisimman aito tai niin, että viestien sisältöä muunnellaan sen tunnistamisen vaikeuttamiseksi

    - tätä on mahdollista myös kohdistaa entistä tarkemmin kohdistetuissa hyökkäyksissä uhria vastaan hänestä kerättyjen tietojen avulla

- mahdollisuus yhdistää eri tekstien sisältöä ja analysoida sekä tunnistaa tätä kautta erilaisia virheitä tai haavoittuvuuksia

    - tämä tarjoaa myös mahdollisuuden julkisten tietojen yhdistelyllä tunnistaa jotain sellaista, joka muodostaa organisaation tai henkilön kannalta kriittistä tai jopa salassa pidettävän tietokokonaisuuden

Koska generatiivinen tekoäly mahdollistaa luonnollisen keskustelun, tämä tarjoaa verkkorikollisille mahdollisuuden luoda esimerkiksi ihmismäiseltä vaikuttava digihuijausbotti, joka kalastelee uhreja verkkopalveluissa, esimerkiksi sosiaalisen median alustoilla tai soittaen puhelimella. Tämä tarjoaa myös erittäin hyvän positiivisen mahdollisuuden luoda erilaisia tukipalveluita, esimerkiksi vanhuksille, mutta samalla luo pelottavan mahdollisuuden väärinkäytöksiin.

Tekoälypalvelut ovat merkittävästi edistäneet tekstin ja myös äänipohjaisten kielikäännöspalveluiden laadun parantumista ja yleistä kehittymistä. Tämän johdosta on odotettavissa, että verkkorikollisten ja muiden väärinkäyttäjien käyttämä teksti parantuu koko ajan ja mahdollisuus sen tunnistaminen esimerkiksi kielioppi- tai

kirjoitusvirheiden johdosta tulee poistumaan.

3.1.5.2 Mahdollisuus tuottaa ja arvioida ohjelmakoodia

Hyökkääjä voi käyttää tätä oman ohjelmakoodin tuottamiseen tai kohteen haavoittuvuuksien etsimiseen

- esimerkiksi haittaohjelma voi mukautua ja muuttaa toimintaansa polymorfisesti, jotta se pysyy havaitsematta mahdollisimman pitkään

- verkkorikollinen voi rakentaa automaattisesti internetissä kulkevan ”botin”, joka hakee ja kirjaa ylös tuntemiaan haavoittuvuuksia ja käynnistää niitä löydettyään, haavoittuvuutta varten tarkoitetun automaattisen hyökkäyksen

3.1.5.3. Syväväärennösten (deepfake) toteuttaminen

Verkkorikolliset voivat jo nyt kohtalaisen helposti kaapata olemassa olevista videoista, äänitiedostoista ja valokuvista materiaalia aidolta näyttävien valokuvien, äänten ja videoiden tuottamiseksi.

Jos nyt varoitamme käyttäjiä siitä, että esimerkiksi jokin perheenjäseneksi tai kaveriksi tekeytyvä verkkorikollinen saattaa lähestyä pikaviestillä pyytäen tietoja, jatkossa tämä sama viesti voi tulla äänitiedostona tai jopa reaaliaikaisena puheena.

Sama koskee videoiden aitoutta, josta olemme nähneet netissä julkaistuja aidolta näyttäviä videoita tunnetuista henkilöistä, jotka ovat kuitenkin osoittautuneet väärennöksiksi. Alkuvuodesta 2024 uutisoitiin verkkorikoksesta Hong Kongissa, jossa yrityksen taloushallinnon asiantuntijaa huijattiin verkkokokouksessa deepfake-videoiden ja hahmojen avulla ja yritys menetti 25 miljoonaa dollaria tässä rikoksessa. [7]

Tämä tekniikka mahdollistaa entistä paremmin informaatiovaikuttamisen ja disinformaation levittämisen sekä myös henkilöiden kiristämisen. Näitä teknologioita tullaan entistä enemmän hyödyntämään esimerkiksi vaalivaikuttamisessa.

3.1.5.4. Esineiden, hahmojen ja kasvojen tunnistaminen ja analysointi

Tekoälypalveluilla voidaan tunnistaa entistä tarkemmin esimerkiksi ihmisten kasvojen ilmeistä ja muusta käyttäytymisestä tunteita, joita voidaan väärinkäyttää heitä vastaan. Tällainen profilointi on EU:n tekoälyasetuksen mukaisesti kiellettyä, mutta valitettavasti asetus ei koske kuin EU-aluetta tai tälle alueelle palveluita tarjoavia toimijoita. Kuten aikaisemmin on todettu, verkkorikolliset eivät noudata lakeja tai muita eettisiä periaatteita.

Jos nyt käytät jotain sovellusta esimerkiksi eläinten tai kasvien, musiikkikappaleiden tunnistamiseen, sama toiminnallisuus voidaan toteuttaa minkä tahansa asian

[7] https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html

tunnistamiseen. Tekoälypalveluiden modaalinen tietojen syöttö mahdollistaa esimerkiksi nyt mahdollisuuden ottaa valokuvan, jatkossa videon, ja pyytää palvelua analysoimaan kuvassa näkyviä asioita, esineitä ja tapahtumia.

3.1.5.5. Tekoälypalveluiden 24/7 kyvykkyys

Kuten kaikki teknologia, myös tekoälypalvelut voidaan valjastaa väsymättä toteuttamaan sille asetettua tehtävää.

Esimerkiksi verkkorikollinen voi laittaa liikkeelle internet-verkkoa läpikäyvän ohjelman, joka kerää ja tunnistaa palveluista löytyviä mahdollisesti salassa pidettäviä tietoja tai palveluihin liittyviä teknisiä haavoittuvuuksia ja käynnistää vaikka kyseiseen organisaatioon kohdistuvan automaattisen kiristysprosessin.

Tällaisia palveluita on ollut käytössä jo pitkään, mutta nyt tekoäly tuo tässä väärinkäyttäjille entistä paremmat mahdollisuudet niiden tunnistamiseksi sekä väärinkäyttämiseksi ja tämän kaiken automatisoimiseksi tekoälyn avulla.

Samoin hyökkääjä voi käyttää tätä kyvykkyyttä nykyistä kehittyneempien ja tarkemmin kohdistettujen palvelunestohyökkäysten toteuttamiseksi

Edellä olevat viisi esimerkkiä ovat osa niistä mahdollisuuksista, joissa verkkorikolliset sekä muut väärinkäyttäjät voivat hyödyntää tekoälyä. Valtaosa menetelmistä on jo osin ollut käytössä pitkään, mutta tulemme kohtaamaan myös uudenlaisia verkkorikollisuuden muotoja sekä olemassa olevien menetelmien merkittävää laadullista kehittymistä.

3.2 Tarkista myös nämä asiat, jos aiot käyttää tekoälypalveluita vapaa-ajalla

Näitä suosituksia voi soveltaa tekoälypalveluiden hankkimiseen myös organisaatioissa.

3.2.1 Miten tuttu ja tunnettu palvelun tarjoaja on?

Markkinoille on tullut ja tulee valtavasti uusia, osin startup-tyyppisiä yrityksiä, joiden palvelut saattavat kärsiä kiireellisestä toteutuksesta ja joiden elinikä voi olla lyhyt. Palvelusta kannattaa hakea mahdollisimman luotettavaa lisätietoa ja kokemuksia, ennen kuin sen ottaa laajempaan käyttöön tai ostaa maksullisen käyttöoikeuden. Useat palvelut tarjoavat aluksi määräaikaisen maksuttoman kokeilujakson, jota kannattaa ehdottomasti hyödyntää ennen palvelun hankintaa. Suosittelemme hyödyntämään tunnetuimpien palveluntarjoajien palveluita, koska heillä on resurssit kehittää ja tuoda niihin uusia ominaisuuksia sekä huolehtia mahdollisten virheiden ja haavoittuvuuksien korjaamisesta pienempiä toimijoita nopeammin.

3.2.2 Varo huijauspalveluita ja tarkista palveluiden käyttöehdot

Myös verkkorikolliset ovat havainneet valtavan kiinnostuksen tekoälypalveluihin. Älä hae palveluita ”googlella” tai lataa palveluita suoraan sinulle tarjotuista linkeistä. Lataa palvelut niiden valmistajien sivuilta ja älylaitteissa Google Play -sovelluskaupasta tai Applen AppStoresta. Myös näissä kauppapaikoissa voi olla sovelluksia, jotka eivät toimi niin kuin ne väittävät. Siksi kannattaa lukea huolella esimerkiksi palveluiden arvostelut.

Varmista myös, ettei palvelun hankkimiseen tai käyttöön liity yllättäviä kuluja tai pidempiaikaisia maksusitoumuksia ennen kuin olet varma siitä, että olet halukas maksamaan palvelun pitempiaikaisesta käyttämisestä kuukausimaksun muodossa.

3.2.3 Tutustu huolella palvelun asetuksiin ja sen käyttösopimuksiin

Tarkista, millaiset oikeudet palvelun tarjoaja saa palveluun vietävään (opetus)dataan sekä miten se hyödyntää palveluun kertyvää tietoa - myös sen käytöstä ja käyttäjistä.

Varmista, onko mahdollista, että opetusdata rajataan vain omaan käyttöösi, jolloin se ei leviä muille palvelun käyttäjille. Tai vaatiiko se esimerkiksi erillisen, maksullisen version hankkimisen.

Varmista, miten voit poistaa palveluun lisätyt tiedot käytön jälkeen. Huomioi, että palveluun esittämäsi kysymykset (kehotteet, prompt) saattavat tulla muiden henkilöiden nähtäville, jos esimerkiksi esittelet palvelua laitteellasi.

Tutustu ja opiskele kuvien, videoiden ja musiikin tuottamiseen tarkoitetuissa palveluissa, miten varmistat, että kyseisiä tuotoksia ei jaeta avoimesti kaikille (public), ellet nimenomaisesti näin halua toimia.

3.2.4 Mieti, mitä sähköpostia tai käyttäjätunnusta käytät palveluiden hyödyntämisessä

Harkitse, pitäisikö tekoälypalveluiden kokeiluja varten luoda uusi sähköpostiosoite. Useat älylaitteet, esimerkiksi Applen iOS, tarjoavat ”Kätke osoitteeni” toiminnon[8]. Silloin käyttäjätunnusta luodessasi sinulle syntyy palvelua varten ainutkertainen sähköpostiosoite, joka ohjaa palvelun lähettämät sähköpostit henkilökohtaiselle sähköpostitilillesi. Saat tätä kautta anonymiteetin kyseiseen palveluun.

Vastaavasti Gmail-sähköpostissa voit lisätä sähköpostiosoitteesi perään + (plus) merkin ja heti perään minkä tahansa merkkijonon. Esimerkiksi aku.ankka@gmail.com ottaa vastaan postia myös osoitteella aku.ankka+tekoaly@gmail.com tai aku.ankka+aika.janna@gmail.com. Samoin voit kokeilla myös käyttää pistettä, koska Gmail-palvelu ei tunnista osoitteissa olevia pisteitä lainkaan, esimerkiksi a.ku.ank.ka@gmail.com menee myös perille.

[8] Kätke osoitteeni -osoitteiden luominen ja hallitseminen iPhonen Asetuksissa - Apple-tuki (FI)

3.2.5 Varmista, että sinulla on oikeudet siihen dataan, jota käytät tekoälyn kouluttamiseen

Oikeuksissa pitää huomioida esimerkiksi tekijänoikeudet sekä luvat tietojen käyttöön ja jakamiseen. Sinulla saattaa työpaikallasi olla käytössä esimerkiksi sellaisia organisaatiota varten tuotettuja tietoja tai raportteja, joiden jakaminen ulkopuolisille tahoille on sopimuksessa kielletty. Tällöin sinulla ei ole oikeutta ladata kyseistä raporttia organisaatiosi sinulle tarjoamaan tekoälypalveluun, jos palvelu hyödyntää kaikkia tietoja sen opetusdatana. Samoin kannattaa varmistaa, voiko tällaista raporttia ladata edes sellaiseen tekoälypalveluun, joka ei tietoja vie opetusdataksi.

4. Esimerkkejä organisaatiolle hallinnollisen tekoälykehyksen luomiseen

Olemme koonneet tähän lukuun muutamia esimerkkejä niistä asioista ja asiakirjoista, joita organisaatioiden kannattaa miettiä laadittavaksi osana sen tekoälypalveluiden hyödyntämisprosessia.

Haluamme korostaa, että organisaatioiden johdon tulee olla nyt **aktiivisesti luomassa ja kehittämässä** organisaatiossa tarvittavaa ”tekoälykehystä”, jonka avulla organisaatio ottaa tekoälypalvelut hallitusti käyttöön. Vaarana on se, että muussa tapauksessa henkilöstö saattaa tuskastua ja hyödyntää vapaa-ajan laitteita, palveluita ja tekoälyä työtehtävien hoitamiseen, siis ns. ”Varjo-AI” -palveluita.

4.1 Henkilöstön osaaminen on keskiössä

ICT-palvelut ja digilaitteet ovat mahdollistaneet meille digipalveluiden hyödyntämisen ja tehtävien suorittamisen uudella tavalla. Tekoälypalveluiden hyödyntäminen ja nopea yleistyminen tulevat jatkossa olemaan osalle käyttäjistä henkilökohtainen osaamishaaste, jota vahvistaa tekoälypalveluiden todella nopea kehittyminen.

Kehityksessä on vaarana, kuten ”ATK:n” yleistyessä 1990- ja 2000-luvulla sekä digitalisaation kehittyessä 2010-luvulla, että osa henkilöistä ei halua opiskella jälleen yhtä uutta teknologiaa tai palvelua. Tai osa käyttäjistä kokee, että uudenlaiset tekoälypalvelut eivät ole heitä varten.

Tekoälypalvelut voivat auttaa kaikkia ikäryhmiä ja on hyvin todennäköistä, että saamme tarjottua uudenlaisia, helppokäyttöisempiä ja intuitiivisempia tekoälypalveluita sekä tekoälyllä varustettuja laitteita esimerkiksi ikäihmisten arkea helpottamaan. Yksi 2020-luvun loppupuolen mahdollinen megatrendi on tekoälypalvelut, joita tullaan todennäköisesti tulevaisuudessa yhdistämään erilaisiin palvelurobotteihin ja metaversumi-ympäristöissä toimiviin avatar-hahmoihin.

Tällä on varmasti vaikutusta siihen, kuinka hankalasti tai helposti pystymme tulevaisuudessa pitämään erossa fyysisen ja virtuaalisen maailman. Näiden kahden eri toimintaympäristön sekoittuminen, keskinäinen vuorovaikutus ja toiminnan ymmärtäminen tulevat merkittävästi nykyistä hankalammin ymmärrettäväksi.

Uusien palveluiden ja laitteiden eräs haaste on tarve oppia pois vanhoista toimintamalleista sekä rohkeasti ottaa käyttöön uudenlaisia palveluita ja laitteita.

Organisaatioiden tulee aktiivisesti kouluttaa henkilöstöään tekoälypalveluiden hyödyntämiseen liittyvistä mahdollisuuksista ja muutoksista. On tärkeää avoimesti kertoa sekä positiivisista kokemuksista että myös ongelmista ja haasteista, joita palveluiden kehittämisessä ja hyödyntämisessä on koettu.

4.2 Tekoälystrategiassa huomioitavia asioita

Tekoälystrategia on organisaation suunnitelma siitä, miten se aikoo hyödyntää tekoälyä tavoitteidensa saavuttamiseksi ja toiminnan kehittämiseksi. Tämä voi käsittää kaikenlaisia asioita tekoälyn käytöstä päivittäisissä toiminnoissa, kuten asiakaspalvelussa, aina suurempiin aloitteisiin esimerkiksi tuotekehityksessä tai uusien liiketoimintamallien luomisessa.

Alla on lueteltu asioita, joihin tekoälystrategiassa tulisi ottaa kantaa:

1. Tavoitteiden ja prioriteettien määrittely

Miksi ja miten tekoälyä halutaan hyödyntää? Mitkä ovat tärkeimmät tavoitteet ja prioriteetit? Tavoitteiden tulisi olla linjassa organisaation laajemman strategian ja tavoitteiden kanssa.

2. Eettiset periaatteet

Hyvä strategia määrittelee organisaation eettiset periaatteet ja säännöt tekoälyn käytölle. Aikaisemmin tässä materiaalissa nostettiin esille mm. seuraavat näkökulmat:

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

3. Tietojen hallinta

Tekoäly tarvitsee laadukasta dataa toimiakseen tehokkaasti ja luotettavasti. Strategiassa tulisi määritellä, miten tietoja kerätään, säilytetään, analysoidaan ja jaetaan organisaatiossa ja sen ulkopuolella. Strategiassa pitää myös määritellä, millaista opetusdataa käytetään tekoälyn kouluttamiseen, jos tekoälyn koulutusprosessi on organisaation hallinnassa, ja kuinka tulosten laatua voidaan mitata ja valvoa. Osaksi tätä prosessia pitää liittää myös opetusdatan laatuun liittyvä seuranta.

4. Osaamisen ja resurssien kehittäminen

Millaista osaamista ja millaisia resursseja organisaatiossa tarvitaan tekoälyn hyödyntämiseksi? Mistä ja miten näitä saadaan sekä miten näitä kompetensseja ja resursseja kehitetään?

5. Teknologian valinta ja käyttöönotto

Minkälaista teknologiaa organisaatio aikoo käyttää, ja miten se otetaan käyttöön? Strategian tulisi sisältää suunnitelma teknologian valinnasta, hankinnasta, testauksesta ja käyttöönotosta. Käyttöönotettavat palvelut tulisi tukea organisaatiossa jo käytössä olevia arkkitehtuuri- tai muita periaatteita.

6. Yhteistyö ja kumppanuudet

Miten organisaatio aikoo tehdä yhteistyötä muiden julkisten organisaatioiden, yksityisen sektorin, yliopistojen ja muiden tahojen kanssa?

7. Mittaaminen ja seuranta

Miten organisaatio aikoo mitata ja seurata tekoälyn käytön vaikutuksia ja tuloksia?

Strategian luomisen jälkeen on tärkeää, että se on elävä dokumentti, jota päivitetään säännöllisesti uusien tietojen, kokemusten ja muuttuvien olosuhteiden mukaan. Lisäksi strategian toteutumista tulee seurata ja tukea riittävillä resursseilla.

4.3 Luonnos tekoälypolitikassa huomioitaviksi asioiksi

Tämä politiikkaluonnos on suunniteltu ohjaamaan organisaatiota tekoälyn käytössä, varmistamaan eettiset ja vastuulliset käytännöt sekä edistämään jatkuvaa parantamista ja oppimista.

1. Strategisten suuntaviivojen asettaminen

Määritämme konkreettisia, mitattavia päämääriä, joiden avulla voimme seurata tekoälyhankkeidemme edistymistä ja arvioida niiden menestystä. Pyrimme aina varmistamaan, että hankkeemme tukevat organisaation yleisiä strategisia tavoitteita.

Esimerkiksi, jos tavoitteemme on parantaa asiakaspalvelua tekoälyn avulla, asetamme konkreettisen tavoitteen, kuinka paljon asiakaspalvelupyyntöjen käsittelyaikaa tulee lyhentää tekoälyn avulla seuraavan vuoden aikana tai kuinka paljon asiakastyytyväisyyden pitää parantua.

2. Vastuullisuus ja avoimuus

Noudatamme eettisiä standardeja kaikessa tekoälytyössämme. Tämä tarkoittaa esimerkiksi syrjimättömyyden, tietosuojan ja avoimuuden periaatteiden noudattamista. Pyrimme myös selittämään tekoälymalliemme toiminnan niin avoimesti kuin se on mahdollista.

Kehitämme käytäntöjä, joiden mukaan kaikki tekoälymallit on testattava syrjinnän varalta ennen käyttöönottoa. Lisäksi pyrimme julkaisemaan malliemme päätöksenteon periaatteet ja algoritmit niin pitkälle kuin se on mahdollista. Huolehdimme oman

opetusdatamme laadun valvonnasta pyrkien sillä minimoimaan tekoälyn tuottamiin tuloksiin liittyvät virheelliset tiedot ja hallusinaatiot, mahdollisen syrjinnän tai vinoumat.

3. Datan hyödyntäminen ja suojaaminen

Keräämme ja hyödynnämme dataa vastuullisesti, noudattaen kaikkia sovellettavia lakeja ja standardeja. Pyrimme myös varmistamaan datan laadun, sillä ymmärrämme, että se on avain tehokkaaseen ja luotettavaan tekoälyn käyttöön.

Säilytämme kaikki keräämämme tiedot turvallisesti ja noudatamme tietosuojalakeja. Lisäksi käytämme dataa tekoälyn kouluttamiseen vain silloin, kun meillä on siihen asianmukaiset oikeudet ja luvat.

4. Henkilöstön koulutus ja osaamisen kehittäminen

Investoimme henkilöstömme tekoälyosaamiseen. Tämä tarkoittaa esimerkiksi koulutusohjelmien järjestämistä, mentorointia tai asiantuntijoiden palkkaamista.

Tarjoamme esimerkiksi vuosittain koulutuksia tekoälystä kaikille työntekijöillemme, ja järjestämme syventäviä työpajoja tekoälyprojekteissa työskenteleville.

5. Sopivimman teknologian valinta

Valitsemme tekoälyteknologiat huolellisesti, ottaen huomioon niiden turvallisuuden, kustannustehokkuuden, käytettävyyden sekä soveltuvuuden organisaatiomme tarpeisiin. Olemme avoimia eri teknologioille ja valitsemme parhaan työkalun kuhunkin tehtävään.

Arvioimme jokaisen tekoälyprojektin alussa, mikä käytössämme oleva tekoälyalusta tai -työkalu on paras kyseiseen tehtävään. Otamme huomioon niin kustannukset, tukiresurssit, yhteensopivuuden muiden järjestelmien kanssa kuin käyttäjäystävällisyyden.

6. Yhteistyöverkoston rakentaminen

Haemme aktiivisesti yhteistyökumppaneita niin julkiselta kuin yksityiseltä sektorilta, mukaan lukien oppilaitokset ja tutkimuslaitokset. Uskomme, että yhdessä voimme saavuttaa enemmän ja luoda parempia tekoälyratkaisuja.

Teemme yhteistyötä sidosryhmien kanssa hyvien tekoälykäytäntöjen kehittämisessä ja jakamisessa.

7. Jatkuva arviointi ja parantaminen

Arvioimme säännöllisesti tekoälyhankkeidemme suorituskykyä ja teemme tarvittavat korjaukset. Käytämme sekä kvantitatiivisia että kvalitatiivisia mittareita, ja otamme

oppia sekä onnistumisista että epäonnistumisista. Pyrimme toimimaan iteratiivisesti, jatkuvasti parantaen ja optimoiden tekoälyn sovelluksia ja prosesseja.

Käytämme esimerkiksi asiakastyytyväisyyskyselyitä ja liiketoimintatietoja arvioidaksemme, miten tekoälymallimme toimivat ja mitä vaikutuksia niillä on ollut. Pidämme myös säännöllisiä palautepalavereita tekoälyprojektitiimien kanssa, jotta voimme oppia suoraan niiltä, jotka työskentelevät tekoälyn parissa päivittäin.

4.4 Riskienhallinnan toteuttaminen

Tekoälyn, kuten muidenkin palveluiden osalta tulee toteuttaa riskienhallintaa. Siinä on suositeltavaa hyödyntää organisaatiossa käytössä olevaa toimintamallia, jota sovitetaan käytettäviin tekoälypalveluihin. Tämä on myös yksi edellytys EU:n tekoälyasetuksen näkökulmasta.

Organisaation kannattaa tunnistaa esimerkiksi tekoälypalveluiden käyttämiseen liittyviä sekä korkean että matalamman tason riskin kohteita:

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

Tekoälysovellukset saattavat tuottaa kovin väärää tietoa, joten riskienhallinnassa pitää ottaa huomioon, miten virheellinen tieto korjataan ja miten tekoälyn tuottama tieto tarkistetaan ennen sen julkistamista.

Osa riskeistä on todennäköisesti sellaisia, joiden hyväksyminen sellaisenaan tai edes lisäkontrollien avulla voi olla organisaatiolle haasteellista. Tämän takia tekoälyn hyödyntämisessä kannattaa etenkin alkuvaiheessa keskittyä sellaisiin tehtäviin, joissa riski on matala tai sitä voidaan hallita. Tällaisia ovat esimerkiksi

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

Yksi keskeisimmistä riskeistä liittyy sellaisen ulkopuolisen palvelutoimittajan tuottaman palvelun käyttämiseen, jonka kanssa ei yleensä ole mahdollista tehdä mitään

sopimusta. Ehdot pitää hyväksyä sellaisenaan, eikä palveluiden turvallisuutta ole mahdollista arvioida.

Yksi keskeinen mahdollisuus hallita riskejä on käyttää julkisen hallinnon käyttöön rakennettua, esimerkiksi kotimaista, Suomessa sijaitsevaa tekoälypalvelua.

Tekoälyyn liittyviä riskiluokkia

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 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. Käytännössä organisaation johdolla on kokonaisvastuu näissä organisaation itse tuottamissa tai sen käyttöön hankkimissa palveluissa.

IPR-riskit (Intellectual Property Rights) eli immateriaalioikeudet

Generatiivisen tekoälyn käyttöön liittyy useita IPR-riskejä, joita organisaatioiden tulisi ottaa huomioon. Jos esimerkiksi tekoälyratkaisu on koulutettu datalla, joka sisältää tekijänoikeuden alaista aineistoa, voi syntyä tekijänoikeuden loukkaus. On tärkeää varmistaa, että koulutusdata on oikein lisensoitu.

On selvää, että tekoäly tulee muuttamaan ja herättämään keskustelua IPR-oikeuksista tulevaisuudessa. On mahdollista, että tarvitaan uusia IPR-järjestelmiä tekoälyn ainutlaatuisten haasteiden käsittelemiseksi.

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.

Riskien hallintakeinoja

Kaikki edellä mainitut riskiluokat tulee käydä läpi osana riskien arviointia sekä kuvata ja toteuttaa niihin riittävät kontrollit jäännösriskin hallitsemiseksi vaadittavalle tasolle. Jos riskiä ei voida pienentää hyväksyttävälle tasolle, niin palvelua ei ole mahdollista toteuttaa tai ottaa käyttöön

Lainsäädännöllinen riskienhallinta

Organisaation on varmistettava, että sen tekoälyratkaisut noudattavat kaikkia soveltuvia lakeja ja asetuksia.

Eettinen riskienhallinta

Organisaation on otettava huomioon eettiset näkökohdat ja varmistettava, että sen tekoälyratkaisut ovat eettisesti kestäviä.

Tietoturvariskien hallinta

Tietoturvastrategioiden ja -käytäntöjen on oltava käytössä riskien hallitsemiseksi.

4.4.1 EU-tekoälyasetus

Tavoitteet ja soveltaminen

’Tekoälyjärjestelmällä’ tarkoitetaan konejärjestelmää, joka on suunniteltu toimimaan eriasteisen itsenäisesti, joka voi mahdollisesti mukautua käyttöönoton jälkeen ja joka voi vastaanottamiensa syöttötietojen perusteella tuottaa tiettyjen eksplisiittisten ja implisiittisten tavoitteiden saavuttamiseksi tuotoksia, kuten ennusteita, sisältöä, suosituksia tai päätöksiä, jotka voivat vaikuttaa todellisiin tai virtuaalisiin ympäristöihin.

Riskiperustainen lähestymistapa

Asetus määrittää neliportaisen asteikon, joista tärkeintä on tunnistaa, mitkä sisältävät ”liian suuren riskin” eli nämä ovat kiellettyjä.

- **Liian suuri riski, kielletyt käytännöt**

  Tähän ensimmäiseen kategoriaan kuuluvat sellaiset tekoälyjärjestelmät, joiden käytöstä aiheutuva riski ei ole hyväksyttävissä. Tähän riskikategoriaan kuuluvien järjestelmien katsotaan aiheuttavan niin selvää uhkaa ihmisten turvallisuudelle, toimeentulolle ja oikeuksille, että niiden käyttö kielletään kokonaan.

  - Subliminaaliset tekniikat, joita henkilö ei havaitse tietoisesti, tai tarkoituksellisesti manipuloivia tai harhaanjohtavia tekniikoita. Tavoitteena on vääristää tai jotka olennaisesti vääristävät henkilön tai henkilöryhmän käyttäytymistä ja päätöksentekoa.
  - Tietyn henkilön tai henkilöryhmän haavoittuvuuksien hyödyntäminen (ikä, vamma, erityistilanne).
  - Sosiaalinen pisteytys, eli henkilöitä tai ryhmiä arvioiva tai luokitteleva järjestelmä, joka johtaa haitalliseen kohteluun, tai kohtuuttomiin seurauksiin.
  - Rikosten ennustaminen henkilön persoonallisuutta koskevilla riskiarvioilla (vrt. Minority report -elokuva) (poikkeuksena, jos arviointi on tukitoiminto ihmisen tekemälle arvioinnille)
  - Kasvojen tunnistaminen haravoimalla kohdentamattomasti kasvokuvia internetistä tai valvontakamerakuvista
  - Työntekijöiden tai opiskelijoiden tunteentunnistus (Poikkeuksena lääketieteelliset ja turvallisuustarkoitukset)
  - Biometrinen tunnistaminen erityisen tiedon perusteella (rotu, uskonto, poliittinen mielipide jne.)
  - Reaaliaikainen biometrinen etätunnistaminen julkisissa tiloissa lainvalvontatarkoituksessa. (Poikkeuksena ihmiskaupan tunnistaminen, terrori-iskun ehkäisy ja rikoksestä epäiltyjen tunnistaminen)

Muut riskiryhmät ovat:

- **Korkea riski, korkeat vaatimukset**
- **Tiettyjen palveluiden rajoitetut riskit, läpinäkyvyysvelvoitteita**
- **Pieni riski, ei vaatimuksia, mutta suosituksia**

Tarkoitus ei ole tähän materiaaliin purkaa tätä kokonaisuutta, vaan antaa käsitys tekoälyasetuksen riskiluokista. Lisätietoa löytyy esimerkiksi täältä Tekoälysäädös |

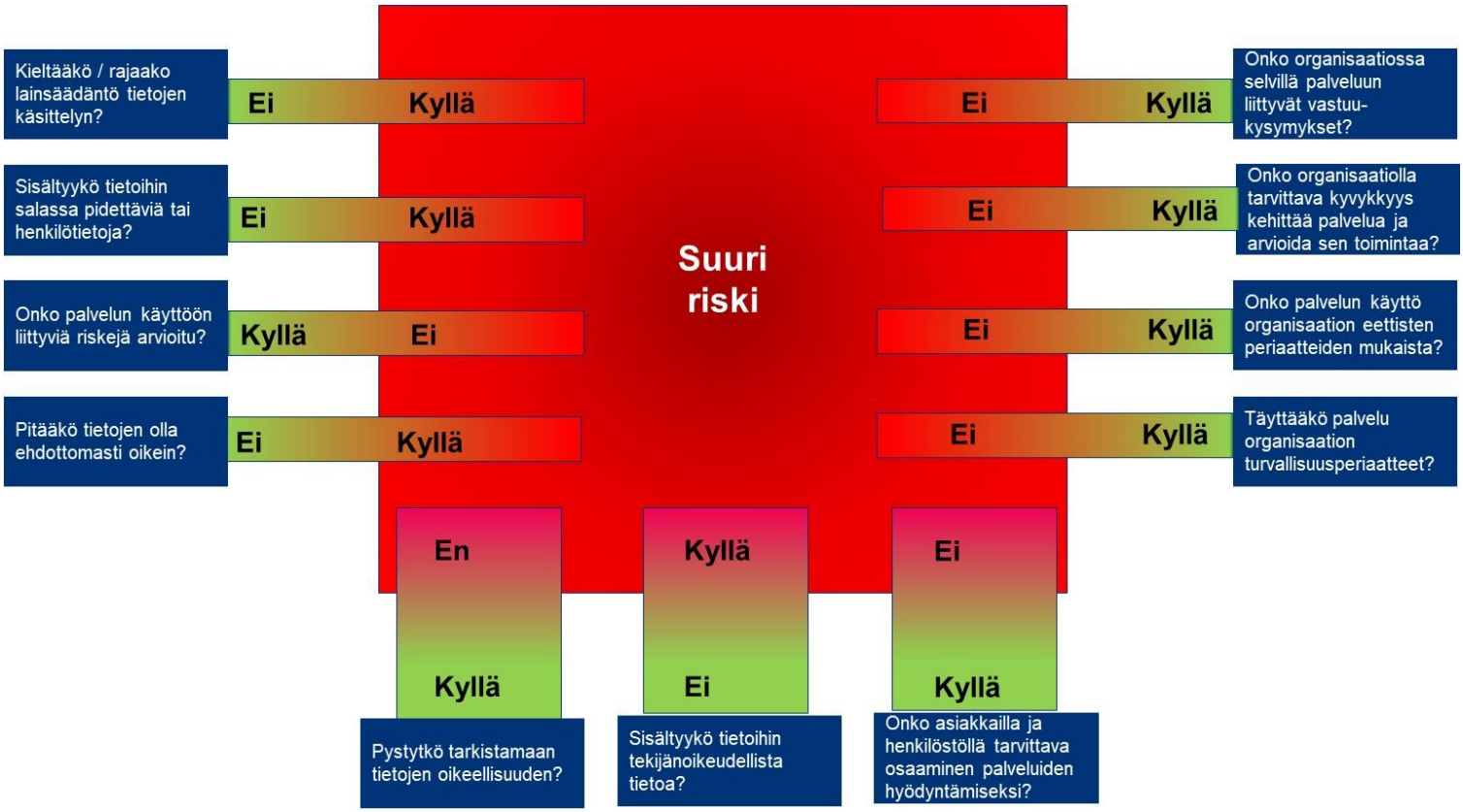
Shaping Europe's digital future (europa.eu) sekä muista tulevista asetukseen liittyvistä tukimateriaaleista. Erityisesti kannattaa tutustua soveltamisaikatauluun, jossa on vaiheita 6 – 9 – 12 – 24 ja 36 päähän sekä seuraamuksia, esimerkiksi varoituksia ja hallinnollisia toimenpiteitä, joista hallinnolliset sakot voivat olla yrityksille todella merkittäviä (jopa 3-7% tai 15 tai 35 miljoonaa euroa yrityksen edeltävän tilikauden vuotuisesta maailmanlaajuisesta kokonaisliikevaihdosta sen mukaan, kumpi näistä määristä on suurempi).

4.4.1 Esimerkki - miten voit helposti arvioida tekoälyn hyödyntämiseen liittyviä riskejä?

Tekoälypalvelun, kuten kaikkien organisaation itse tuottamien tai käyttöönottamien palveluiden, käyttö edellyttää jatkuvaa riskienhallintaa.

Käy läpi esitetyt kysymykset ja vastaa niihin. Mitä enemmän vastauksia sijoittuu punaisella (suuremman riskin) alueella, sitä enemmän tulee varmistaa, että palvelun käyttöönottaminen tai hyödyntäminen on mahdollista, laillista ja turvallista sekä organisaation linjausten mukaista.

Tätä samaa riskiallasmallia voidaan hyvin soveltaa myös muiden kuin tekoälyyn liittyvien palveluiden suunnittelussa ja arvioinnissa.



5. Lainsäädäntö ja tekoäly

Ota huomioon tekoälyyn käyttöön liittyvät yleislait sekä tunnista organisaatiosi kohdistuva erityislainsäädäntö, jolla voi olla tekoälyn käyttöön liittyvää vaikutusta. Alle on koottu muutamia esimerkkejä, tämä lista ei ole täydellinen.

5.1.1 Julkisuuslaki

Laki viranomaisten toiminnan julkisuudesta 621/1999 - Ajantasainen lainsäädäntö - FINLEX ®

22§ Asiakirjasalaisuus

”Salassa pidettävää viranomaisen asiakirjaa tai sen kopiota tai tulostetta siitä ei saa näyttää eikä luovuttaa sivulliselle eikä antaa sitä teknisen käyttöyhteyden avulla tai muulla tavalla sivullisen nähtäväksi tai käytettäväksi.”

5.1.2 EU:n yleinen tietosuoja-asetus (GDPR) sekä kansallinen lainsäädäntö:

EUROOPAN PARLAMENTIN JA NEUVOSTON ASETUS (EU) 2016/ 679, - annettu 27 päivänä huhtikuuta 2016, - luonnollisten henkilöiden suojelusta henkilötietojen käsittelyssä sekä näiden tietojen vapaasta liikkuvuudesta ja direktiivin 95/ 46/ EY kumoamisesta (yleinen tietosuoja-asetus) (europa.eu)

i. käsittelyn lainmukaisuus (6 artikla)
ii. läpinäkyvä informointi, viestintä ja yksityiskohtaiset säännöt rekisteröidyn oikeuksien käyttöä (12 artikla)
iii. tietojen oikaiseminen ja poistaminen (16–17 artiklat)
iv. oikeus käsittelyn rajoittamiseen (18 artikla)
v. sisäänrakennettu ja oletusarvoinen tietosuoja (25 artikla)
vi. tietosuojaa koskeva vaikutustenarviointi (35 artikla)
vii. ennakkokuuleminen (36 artikla)

5.1.3 Tiedonhallintalaki

Laki julkisen hallinnon tiedonhallinnasta 906/2019 - Säädökset alkuperäisinä - FINLEX ®

15 § Tietoaineistojen turvallisuuden varmistaminen

Viranomaisen on varmistettava tarpeellisin tietoturvallisuustoimenpitein, että sen:

1) tietoaineistojen muuttumattomuus on riittävästi varmistettu;
2) tietoaineistot on suojattu teknisiltä ja fyysisiltä vahingoilta;

3) tietoaineistojen alkuperäisyys, ajantasaisuus ja virheettömyys on varmistettu;

4) tietoaineistojen saatavuus ja käyttökelpoisuus on varmistettu;

5) tietoaineistojen saatavuutta rajoitetaan vain, jos tiedonsaantia tai käsittelyoikeuksia on laissa erikseen rajoitettu;

6) tietoaineistot voidaan tarvittavilta osin arkistoida.

5.1.4 Hallintolaki – 2. luku Hyvän hallinnon perusteet

Hallintolaki 434/2003 - Ajantasainen lainsäädäntö - FINLEX ®

- yhdenvertaisuusperiaate (Hallintolaki 6 §)
- tarkoitussidonnaisuuden periaate (Hallintolaki 6 §)
- objektiviteettiperiaate (Hallintolaki 6 §)
- suhteellisuusperiaate (Hallintolaki 6 §)
- luottamuksensuojaperiaate (Hallintolaki 6 §)

5.1.5 Ehdotus Euroopan parlamentin ja neuvoston asetus tekoälyä koskevista yhdenmukaistetuista säännöistä (tekoälysäädös) ja tiettyjen unionin säädösten muuttamisesta

Uusin 21.5.2024 hyväksytty versio:
https://data.consilium.europa.eu/doc/document/PE-24-2024-INIT/en/pdf

Vanha 21.4.2021 versio

https://eur-lex.europa.eu/legal-content/FI/TXT/?uri=celex%3A52021PC0206

5.1.6 Tekoälypalveluiden (kyber) (tieto) turvallisuudelta edellytettävät vaatimukset

Huomaa, valtaosa tekoälypalveluista hyödyntää globaaleja pilvipalveluita. Tällöin näiden käyttämisessä ja hankkimisessa pitää varmistaa palveluiden turvallinen ja luotettava käyttö. Organisaation tulee varmistua, että käytettävät pilvipalvelut täyttävät organisaatioon lainsäädännöstä tulevat vaatimukset koskien ICT-palveluiden hyödyntämistä ja hankkimista. Aikaisemmin mainitun tiedonhallintalain lisäksi organisaation on tunnistanut siihen kohdistuvan erityislainsäädännön sekä 18.10.2024 sovellettavaksi tulevan kyberturvallisuuslain (HE 57/2014)

Kyberturvallisuuslaki
https://www.eduskunta.fi/FI/vaski/HallituksenEsitys/Sivut/HE_57+2024.aspx

Liite 1 – Esimerkkejä tekoälyn soveltamisesta

Tekoälypalvelut kehittyvät tällä hetkellä erittäin nopeasti. Älä tuomitse tai aliarvioi niiden toimintaa ja merkitystä yksittäisten virheiden tai epätarkkuuksien takia. Tässä kappaleessa on kuvattu ensisijaisesti vuosien 2023-2024 aikana nopeasti kehittyneitä uudenlaisia käyttömahdollisuuksia. Tekoälypalveluiden käyttäminen tulee kiinnostamaan yhä enemmän, koska tekoälypalvelut pystyvät luomaan uutta sisältöä tehokkaasti ja auttamaan meitä arkisissa askareissa. Seuraavissa kappaleissa on kuvattu mahdollisia käyttötarkoituksia.

1.1 Palvelut voivat käsitellä valtavan määrän tekstiä tehden niistä tiivistelmiä tai yhteenvetoja sekä vastata sisältöön liittyviin kysymyksiin

Eikö olisikin houkuttelevaa pyytää tekoälypalvelua ”tuota edellisen kokouksen nauhoitteesta pöytäkirja” tai kysyä ”mitä olivat meidän edellisen kokouksen päätökset?” tai ”kuka kertoi viime vuoden loppupuolen kokouksessamme ensimmäisen kerran OpenAI tekoälypalvelusta?”

Organisaation tulee ymmärtää, mitä tällaisten toimintojen hyödyntäminen tarkoittaa esimerkiksi tietojen luottamuksellisuuden, eheyden ja saatavuuden sekä henkilötietojen käsittelyn osalta. Lisäksi tekoälypalveluiden hyödyntämiseen liittyy muita, kokonaan uudenlaisia vaatimuksia koskien niiden tuottaman tiedon tai päätösten oikeellisuutta, läpinäkyvyyttä ja eettisyyttä.

1.2 Palvelut voivat kääntää tekstiä, puhetta tai lukea kuvien sisältämää tekstiä ja kääntää sitä kielestä toiseen

Markkinoilla on ollut pitkään lukuisia kehittyneitä, kielestä toiseen kääntäviä perinteisiä ja nyt myös koneoppimiseen tai kielimalleihin pohjautuvia käännöspalveluita, joiden laatu koko ajan paranee. Näiden palveluiden käyttäminen työtehtäviin hoitamiseen tulee ohjeistaa, kuten kaikkien muidenkin palveluiden.

Suositeltavaa on, että esimerkiksi organisaatio hankkii lisenssin turvalliseksi arvioimaansa palveluun ja ohjeistaa, millaista tietoa kyseisen palvelun avulla saa kääntää.

Myöskään vapaa-ajalla käännöspalveluihin ei kannata syöttää mitään salassa pidettäviä, henkilökohtaisia tai muuten yksilöiviä tietoja, ellet voi varmistua palvelun turvallisuudesta.

Uusimmat palvelut tulevat keskustelemaan kanssasi luontaisesti ihmisenkaltaisesti, jopa vaihtaen tarvittaessa ääntä tunteen mukaisesti tai suoraan kielestä toiseen.

1.3 Palvelut voivat luoda ja tuottaa artikkeleita, sähköpostiviestejä sekä melkein mitä tahansa sisältöä

Tekoälypalvelut ovat erittäin nopeita ja tehokkaita luomaan ja tuottamaan uusia tekstisisältöjä pohjautuen siihen valtavaan tietomäärään, jolla ne on opetettu. Opetusdatan merkitys korostuu, koska se vaikuttaa palvelun tuottamaan sisältöön ja

vääristymät sekä epätarkkuudet voivat ohjata tuotettua tietoa käyttäjän huomaamatta virheelliseen, väärään suuntaan.

Organisaation tulee ymmärtää, millaisia uhkia palveluiden käyttämiseen liittyy sekä ohjeistaa ja neuvoa, kuinka palveluiden tuottaman tiedon oikeellisuus tulee varmistaa.

Hallusinaation tai muiden virheellisyyksien tunnistaminen esimerkiksi kuvissa tai videoissa, musiikissa voi olla haastavampaa kuin esimerkiksi tekstitulosten osalta.

1.4 Palvelut voivat luoda musiikkia, kuvia ja videosisältöä

Tekoälypalvelut tuovat merkittäviä uusia mahdollisuuksia sellaisille henkilöille, joilla ei ole aikaisemmin ollut mahdollisuutta tuottaa tällaista sisältöä. Milloin saamme kuultavaksemme ensimmäisen globaalin tekoälyn luoman hittikappaleen ja milloin myös sen esittäjänä toimii tekoälyn luoma hahmo?

Tekstin, kuten muunkin tuotetun materiaalin, osalta tulee huomioida mahdolliset tekijänoikeuskysymykset.

Verkkorikollisille ja muille väärinkäyttäjille tekoäly tarjoaa mahdollisuuden luoda kokonaan uudenlaisia digihuijauksia. Jos tähän saakka erilaiset syväväärennykset (puheääni, videot, valokuvat) ovat olleet hitaita valmistaa ja vaatineet hieman kehittyneempää osaamista, uudet tekoälypalvelut tulevat tekemään tästä helppoa kenelle tahansa, hieman vaivaa näkevälle.

Jatkossa informaatioturvallisuuteen ja tietojen oikeellisuuteen liittyvästä osaamisesta tulee entistä tärkeämpi kansalaistaito. Todennäköisesti joudumme jatkossa yhä enemmän miettimään sitä, miten voimme vahvistaa joidenkin julkisten kriittisten tietojen, esimerkiksi julkaistun videon, aitouden esimerkiksi siihen liitetyllä varmenteella.

1.5 Työtehtävien muuttuminen – esimerkkinä ohjelmointi

Useissa yhteyksissä on nostettu esille ammatteja, joihin uusien tekoälypalveluiden arvioidaan vaikuttavan. Esimerkkejä muutoksesta erilaisiin ammatteihin löytyy esimerkiksi asiakirjasta ”GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models”[9].Esimerkkeinä on nostettu esille sovelluskehitys ja ohjelmointi.

Jokainen teknologinen vallankumous, millaisena tätä nopeasti edistyvää tekoälypalveluiden kehittymistä voidaan myös pitää, on vaikuttanut laajasti työn tekemiseen. Alkuvaiheessa uudet tekoälypalvelut tuovat useille toimialoille ja ammatteihin lisää tuottavuutta, kun sen avulla voidaan tehdä ”lyhyessä ajassa enemmän.” Sama pätee esimerkiksi sovelluskehitykseen, mutta kuten jo edellä on nostettu esille, organisaation tulee luoda selkeä ohjeistus ja politiikka tekoälypalveluiden hyödyntämiseen sovelluskehityksessä, sekä omassa organisaatiossa että alihankintaverkostossa.

[9] 2303.10130.pdf (arxiv.org)

Tekoälyn käyttäminen sovelluskehityksessä voi virheellisesti aiheuttaa organisaation salassa pidettävien tietojen, ohjelmakoodin, dokumentaation, algoritmien sekä tuotanto- tai testausdatan päätymisen väärin käsiin. Jonkun pitää myös tarkistaa tuotetun koodin toimivuus. Vaikka koodi saattaa näyttää toimivalta, onko se muuten laadukasta, tehokasta ja organisaation tarkoitukseen sopivaa? Tekoälyn hyödyntäminen saattaa heikentää organisaation omaa kyvykkyyttä ja osaamista sekä luoda uudenlaisia riippuvuuksia tekoälypalveluihin.

Koska tekoälypalvelut pohjautuvat koulutusdataan, sen vinoutumat ja virheet voivat heijastua myös tuotettuun koodiin.

Tulemme jatkossa tarvitsemaan myös uudenlaisia ammattilaisia, jotka pystyvät arvioimaan esimerkiksi erilaisten algoritmien tai palvelun tuottamien päätösten oikeellisuutta. Jos palveluiden tietoturvallisuuden arvioinnin merkitys on nyt tärkeää, jatkossa tarvitsemme asiantuntemusta myös toimintalogiikan ja palvelun tuottamien tietojen tai päätösten arviointiin ja jäljitykseen.

1.6 Tekoälyn kytkeminen olemassa oleviin palveluihin API-rajapintojen avulla

Keskusteleva, vuorovaikutteinen tekoälypalvelu tuo sinällään jo monelle käyttäjälle uudenlaisia mahdollisuuksia sen hyödyntämiseen. Toinen merkittävä, edelleen kehittyvä mahdollisuus on tekoälyn kytkeminen API-rajapintojen kautta olemassa oleviin palveluihin. Silloin käyttäjä voi käyttää tekoälypalvelun käyttöliittymää myös siihen kytketyn palvelun tietojen hyödyntämiseen. Pian kaikki palvelut ovat tekoälyn hyödynnettävissä, jos palvelun tarjoaja haluaa sen mahdollistaa.

Eikö olisi kätevää pyytää hakemaan edullisin lento tai laivamatka halutulle aikavälille paikasta A paikkaan B, ehdotuksia neljän tähden hotelleista kyseiselle aikavälille määrätyllä alueella, suosittelemaan autovuokraamoa tai tekemään matkaohjelma aikatauluineen kohteen nähtävyyksistä vaikkapa viikon matkaa varten sekä suosittelemaan paikallisia, kohtuuhintaisia hyvän arvion saaneita lounasravintoloita?

Tämä kaikki onnistuu tälläkin hetkellä, mutta useamman eri palvelun ja hakutoiminnon tuloksena, johon voi kulua merkittävä määrä aikaa. Edellä olevan tehtävän suorittaminen onnistuu tekoälypalveluiden avulla jo tällä hetkellä käytettäessä sopivia matkustuspalveluihin liittyviä API-rajapintoja.