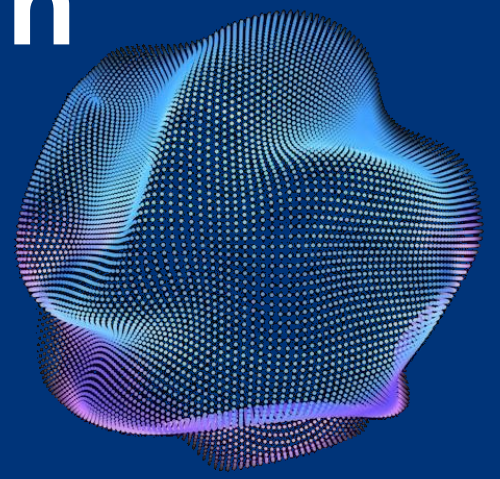
VAHTI-webinaari: Tekoälyn ohjeistaminen ja VAHTI-hyvät käytännöt tukimateriaalin julkistaminen



12.9.2023 Kaj Mustikkamäki, Valtiokonttori ja Kimmo Rousku DVV

DIGI- JA VÄESTÖTIETO-VIRASTO

VAHTI
Tekoälyn huoneentaulut

Tiivistelmä

DIGI- JA VÄESTÖTIETO-VIRASTO

Huoneentaulu johdolle

- **Luo tekoälystrategia**
  - Kaikille selväksi, mihin tekoälyn käytöllä organisaatiossa pyritään
- **Luo tekoälypolitiikka**
  - Miten strategian tavoitteisiin päästään
  - Mihin ja miten tekoälyä saa ja voi käyttää
  - Miten organisaatiossa käytetään tekoälyä
- **Ylläpidä ja uudista tekoälystrategiaa ja –politiikkaa**
  - Lainsäädäntö, teknologinen kehitys, tekoälyn hyötyjen kehittymisen myötä
- **Varmista, että tekoälyn hyödyntämisessä tunnistetut riskit käsitellään ja otetaan hallintaan**
  - teknologian ja järjestelmien kyvykkyyksien muuttuessa
  - lainsäädännön ja muun regulaation kehittymisen mukaan
  - tekoälyn käytöstä tunnistettujen hyötyjen ja haittojen perusteella.
- **Varmista, että organisaatiolla on osaavat asiantuntijat, jos käytätte tekoälyä**
- **Varmista sopimusten vastuut**
- **Varmista, että henkilökunta tuntee strategian ja linjaukset ja on ohjeistettu tekoälyn käytöstä**

Huoneentaulu asiantuntijoille

- Varmista, että tekoälyn käyttö ja kehittäminen, mahdolliset ohjeistukset sekä lisenssiehdot **vastaavat organisaation strategiaa ja politiikkaa**
- Huomioi tekoälyn käytön yhteydessä mahdolliset **tekijänoikeudelliset kysymykset**
- Ohjeista selkeästi henkilöstöä **sallittujen ja kiellettyjen aineistojen käytöstä**
  - Esimerkkejä kielletyistä voivat olla henkilötiedot, salassa pidettävät tiedot, muut tekijänoikeudelliset tiedot jne.
  - Täysin julkinen tieto voi olla sallittua.
- Varmista ja sovi kirjallisesti missä määrin ja millä edellytyksillä tekoälypalveluita voidaan mahdollisesti **hyödyntää hankittaessa alihankintana sovelluskehitystä tai muita palveluita**
- **Ohjeista ja kouluta henkilöstöä** tekoälyyn liittyvistä periaatteista, muun muassa
  - Avoin ja läpinäkyvä käyttö, Oikeudenmukaisuus ja syrjimättömyys, Tietosuoja ja yksityisyys, Turvallisuus ja luotettavuus, Ihmiskeskeisyys, Eettinen suunnittelu ja käyttö, Jatkuva käyttöpalautteen ja virhetietojen käsittely sekä poikkeamatilanteet Tekoälyn käytöstä viestiminen, Osallistuminen ja yhteistyö
- Huomioi tekoälyn hyödyntämiseen liittyvien **eettisten ohjeiden ja riskienhallinnan kehittyminen** sekä päivitä organisaation ohjeistuksia ja prosesseja vastaavasti.
- Tekoälyn käytön alkuvaiheessa arvioi tilannetta tapauskohtaisesti. Yleisesti pätevät ohjeet muokkautuvat kokemuksen karttuessa.
  - Oleellista on kerätä tietoa tekoälyn hyödyntämisen ja soveltamisen osalta

Huoneentaulu henkilöstölle

- **Noudata** organisaation **tekoälypolitiikkaa ja ohjeistuksia** tekoälyn käytöstä
- **Käytä** vain **organisaation hyväksymiä** tekoälysovellutuksia
    - Mikäli käyttämäsi ohjelmisto tai verkkopalvelu ilmoittaa ottaneensa käyttöön tekoälyyn perustuvia ominaisuuksia, ilmoita tästä ohjelmistojen hyväksymisistä vastaavalle taholle
- Käytä tekoälysovellutuksia vain **organisaation hyväksymään tarkoitukseen**
- **Ilmoita** sovitulla tavalla tekoälyn tekemistä **erityyppisistä virheistä**. Pienetkin virheet voivat kertautua huomattavaksi ongelmaksi
- Mikäli tekoäly tuottaa huonolaatuista, virheellistä, vaarallista tai muuten epäilyttävää ja sopimatonta materiaalia, **lopeta sen käyttö**. Dokumentoi kopioimalla tekstit, ottamalla kuvakaappauksia tai tallentamalla mahdolliset virhe- ja lokitiedot analysointia varten
- Varmista aina tekoälysovellutusta käyttäessäsi, ettet syötä sovellutukselle organisaation **kieltämää materiaalia**. Jos olet epävarma, **kysy organisaation tekoälyvastaavalta**
- **Pidä itseäsi ajan tasalla**, osallistu koulutuksiin ja muihin tilaisuuksiin
- **Anna palautetta** tekoälyvastaavalle, jos huomaat ohjeissa ”porsaanreiän” tai muun epäkohdan
- Noudata hyvän, **avoimen hallinnon periaatteita** ja muista **virkavastuu**

Tarkistuslista käyttäjälle ja kehittäjälle, poimintoja listalta ”Tee näin”

- Käytä työnantajan käyttöösi antamia tekoälypalveluita **annettujen ohjeiden mukaisesti**
- **Ymmärrä tekoälyn rajoitukset**. Jokainen tekoälyjärjestelmä on suunniteltu tiettyyn tarkoitukseen ja sillä on omat rajoitteensa
- **Tarkista**, onko käyttämäsi palvelu kaikille avoin, yleinen tekoälypalvelu vai oman organisaatiosi tarjoama palvelu ja **noudata niiden käytöstä annettuja ohjeita**
- Suunnittele tekoälyjärjestelmä **ihmiskeskeisesti**. Siten, että se on helppokäyttöinen ja palvelee käyttäjiensä tarpeita. Tekoälyjärjestelmän tulisi **aina palvella ihmisen tarpeita**, ei päinvastoin
- Varmista, että palvelun tuottamat ratkaisut ovat **eettisesti hyväksyttäviä, syrjimättömiä ja noudattavat yksityisyydensuojaa**
- **Kouluta henkilöstöä**, jotta kaikilla työntekijöillä on tarvittava tieto ja koulutus tekoälyn turvalliseen käyttöön
- **Arvioi ja testaa tekoälyjärjestelmien turvallisuus**
- **Tee ilmoitus** kaikista palveluiden käyttöön liittyvistä ongelmista, tunnistamistasi uhkista ja havaitsemistasi riskeistä organisaatiosi antamien ohjeiden mukaisesti.
- **Varaudu mahdollisiin ongelmiin**. Tämä voi tarkoittaa esimerkiksi varasuunnitelmien tekemistä ja nopeaa viestintää.
- **Kerro**, jos olet hyödyntänyt **materiaalien tuottamisessa** tekoälypalveluita.
- **Ylläpidä avointa viestintää**. Olipa kyseessä sisäinen tiimi tai ulkoiset sidosryhmät, avoin ja jatkuva viestintä on avain tekoälyn turvalliseen ja tehokkaaseen hyödyntämiseen.
- Näe tekoäly **positiivisena**, uudenlaisena, osaamistasi laajentavana **tukipalveluna**.

Tarkistuslista käyttäjälle ja kehittäjälle, poimintoja listalta ”Älä tee näin”

- Älä käytä tekoälyä ilman tarkoitusta. Se on vain työkalu, jonka käytön pitäisi aina palvella selkeää tarkoitusta.
- Älä unohda ihmistä tekoälyn takana. Tekoälyjärjestelmät ovat ihmisten suunnittelemia ja toteuttamia, joten huomioi myös heidän näkökulmansa ja tavoitteensa.
- Älä oleta, että tekoäly ratkaisee kaikki ongelmasi. Tekoäly voi auttaa monissa asioissa, mutta se ei ole ratkaisu kaikkeen. Älä oleta, että tekoäly voi korvata kaikki muut työkalut ja prosessit.
- Älä syötä julkiseen palveluun mitään salassa pidettäviä tietoja tai henkilötietoja.
- Älä syötä julkiseen palveluun sellaista tietoa, johon sinulla tai organisaatiolla ei ole tekijänoikeutta.
- Älä oleta, että tekoäly on aina oikeassa. Tekoäly ei ole erehtymätön, ja sen päätökset voivat olla vääriä. Älä luota sokeasti sen päätöksiin tai tuottamaan tietoon. Tarkasta tekoälyn tuottamat tiedot muuta kautta, jos mahdollista.
- Älä unohda jatkuvaa seurantaa. Tekoälyn turvallisuus ei ole kertaluonteinen tapahtuma, vaan jatkuva prosessi, joka vaatii säännöllistä seurantaa ja päivitystä.
- Älä sivuuta käyttäjien palautetta. Käyttäjien palaute on arvokas resurssi tekoälyn turvallisuuden ja tehokkuuden parantamisessa. Älä jätä huomiotta heidän kokemuksiaan ja ehdotuksiaan.

Muuta huomioitavaa

- Tekoäly luo valtavia mahdollisuuksia, mutta myös merkittäviä riskejä. Verkkorikolliset, valtiollisen tason rikolliset toimijat ja muut tahot pystyvät hyödyntämään sitä helposti välittämättä lainsäädännöstä tai toiminnan eettisyydestä.
- Tekoälyn hyödyntäminen tulee merkittävästi laajentumaan siitä, miten kuvittelemme sitä käytettävän tänä päivänä.
- Tekoälyn käytöstä on useita hyötyjä, mutta sen vastapainona se myös kuluttaa resursseja ja lisää riskejä. Vastuullisessa tekoälyn hyödyntämisessä huomioidaan myös, onko tekoälyn käyttö kaikissa käyttötapauksissa oikeasuhtainen valinta.
- Käytä tekoälyä silloin, kun onnistuminen on todella hyödyksi, mutta epäonnistuminen ei juurikaan tuota vahinkoa.
- Tekoälyllä ei ole suunnitelmaa tai pyrkimystä oikeaan tulokseen, väärin alkanut polku päätelmiä voi johtaa yhä vain huonompaan tulokseen.
- Tekoäly antaa tilastollisesti oikean vastauksen, ei oikeaa vastausta; sen tuottama vastaus on niin hyvä tai kun huono kuin on sen opetuksessa käytetty data.
- Tekoäly on tehokas tuottamaan vastauksia, jolloin virheet monistuvat nopeasti ja tämä tulee huomioida kehitettäessä tekoälyä hyödyntävää palvelua tai prosessia.
- Tulevaisuudessa opetusdatan manipulointi tulee olemaan kasvava riski, joten tekoälyn tuottaman tiedon laatua tulee valvoa.

Entä verkkorikollisuus?

- **rajaton mahdollisuus analysoida ja tuottaa tekstiä eri kielillä laadukkaasti**
- **mahdollisuus tuottaa ja arvioida ohjelmakoodia**
- **syväväärennösten (deepfake) toteuttaminen**
- **esineiden, hahmojen ja kasvojen tunnistaminen ja analysointi**
- **tekoälypalveluiden 24/7 kyvykkyys**

DIGI- JA VÄESTÖTIETOVIRASTO

dvv.fi