

Pieni jalanjälki, suuri kädenjälki – vastuullisen datanhallinnan hyödyt

Dataintensiivisessä ajassamme esimerkiksi työsuoritteet, puhelimen sijaintitiedot ja älysormukset tuottavat informaatiota, jota voidaan käyttää yksilöiden seurantaan ja kehittämiseen. Kun datasta jalostetaan tietoa, voivat organisaatiot tehdä parempia päätöksiä ja tarjota asiakkailleen yksilöllisempiä palveluita. Datan käsittelyn ja hallinnan yhteydessä pitäisi hyötyjen lisäksi tarkastella myös eettisiä ulottuvuuksia ja esimerkiksi tekoälyn eettisiä riskejä.

TEKSTI: MIKKO ELOHOLMA

Organisaatiot ovat jäsentäneet vastuullisuustyötään pitkään ESG-viitekehyksen kautta, jossa ympäristö (Environment), yhteiskuntavastuu (Social) ja hyvän hallintotapa (Governance) sisältävät erilaisia vastuullisuuden mittareita. Data kytkeytyy jokaiseen näistä osa-alueista. Ympäristö kuormittuu ICT-laitteista ja datan käsittelyllä on hiilijalanjälki, mutta toisaalta digitalisaatiolla voidaan tukea esimerkiksi kiertotalouden toteutumista. Algoritmit saattavat huomaamatta asettaa ihmisryhmiä eriarvoiseen asemaan, mutta toisaalta tekoälyllä voidaan saada uutta dataa yhteiskunnallisista ilmiöistä, ja siten tehdä oikeudenmukaisempia päätöksiä. Datan hyödyntämisellä voi siis olla sekä negatiiviseen jalanjälkeen että positiiviseen kädenjälkeen liittyviä eettisiä vaikutuksia.

Suuri osa dataetiikan kysymyksistä koskettaa laajemmin kaikenlaisen tiedon etiikkaa. Dataa hyödynnettäessä sekä positiiviset että negatiiviset vaikutukset voivat kuitenkin moninkertaistua, koska informaatiota käsitellään valtavia määriä osana

Hyödyntääkö organisaatio tietoa ja dataa vastuullisesti?

nopeita, monimutkaisia ja automaattisia prosesseja. Datavastuullisuutta voi tarkastella laajasti kysymällä, hyödyntääkö organisaatio tietoa ja dataa vastuullisesti, eli viime kädessä eettisesti. Tähän kysymykseen vastaaminen vaatii lähestymistä monista näkökulmista.

Datavastuullisuuden näkökulmia

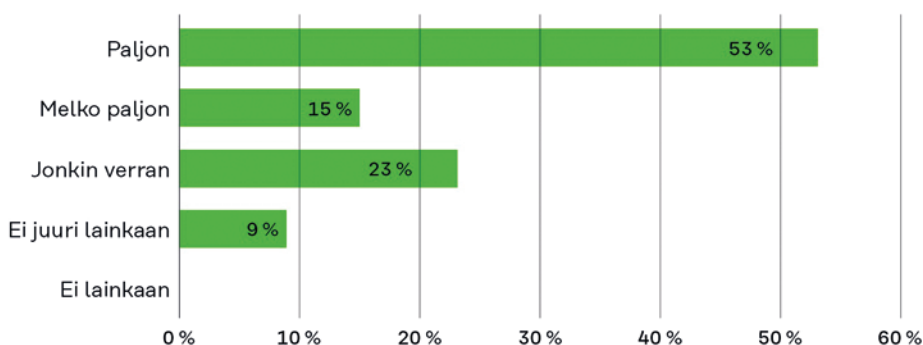
Yhteiskunnan lainsäädäntö pyrkii mukailemaan vallitsevaa eettistä konsensusta, ja **lainsäädännön noudattamista** voi pitää siksi myös datavastuullisuuden kulmakivenä. On tilanteita, joissa organisaatio saattaa huomaamattaankin hyödyntää dataa esimerkiksi tietosuojalainsäädännön, tekijänoikeuslainsäädännön tai julkisen hallinnon tiedonhallintalain vastaisesti.

Datan todenmukaisuus ja laadukkuus on ennakkoehto kaikelle mielekkäälle päätöksenteolle.

Data ei koskaan täysin vastaa todellisuutta, mutta jos se ei heijastele sitä riittävällä tasolla, eivät päätöksetkään yleensä johda hyvin lopputuloksiin. Todenmukaisuutta täytyy pohtia myös datan edustavuuden näkökulmasta: vaikka yksittäisen datapisteet pitäisivätkin paikkansa, voi tietoa-aineisto olla liian vinoutunutta tai rajoittunutta haluttujen johtopäätösten tekemiseksi.

Datan yksipuolisuudesta ja vinoutuneisuudesta puhutaan paljon tekoälyn etiikan yhteydessä. Eräässä tunnetussa esimerkissä Amazon (2018) halusi hyödyntää tekoälyä rekrytoinnissaan, mutta

Datan vastuullinen käyttö tuo yrityksen liiketoiminnalle lisäarvoa oman näkemykseni mukaan...



Sitran (2020) selvityksessä suurin osa vastanneista yritysedustajista näki, että datan vastuullinen käyttö tuo yrityksen liiketoiminnalle lisäarvoa paljon tai melko paljon. Kuva: Sitra.



Open Data Institutin Data Ethics Canvas on työkalu, jolla voidaan tunnistaa projektin, tuotteen tai palvelun eettisiä vaikutuksia, ja sopia tarvittavista toimenpiteistä. Canvasista hyödynnetään aktiivisesti TIEKEN Datavastuullisuuden valmennuksessa. Kuva: Open Data Institute.

huomasi tekoälyn lopulta järjestelmällisesti syrjivän naisten työhakemuksia. Tämä johtui siitä, että lähdedatana oli käytetty yhtiön saamia aiempia työhakemuksia, joissa miehet olivat selvästi yliedustettuina. Lähdedatan vinoumat periytyvät herkästi tehtyihin johtopäätöksiin niin tekoälyn kuin ihmisenkin päätelyssä.

Datan hyödyntämisen seurauksia kannattaa siis pohtia **ihmisryhmien ja heidän yhdenvertaisuutensa näkökulmasta**. Suomessa yhdenvertaisuusvaltuutettu teki 2018 ratkaisun, jossa se katsoi yhtiön evänneen henkilöltä luottoa syrjivällä tilastollisella pisteytysjärjestelmällä, joka hyödynsi dataa henkilön asuinpaikasta, sukupuolesta, iästä ja äidinkielestä. Ratkaisun mukaan tietojen kysyminen ei sinänsä ollut ongelma, mutta tietoja hyödynnettiin arvioinnissa syrjivällä tavalla.

Datan hyödyntämisen riskit voivat liittyä **yksilön oikeuksiin ja yksityisyyteen** esimerkiksi silloin, jos tietoa ei ole onnistuttu anonymisoimaan tai sitä

Datan todenmukaisuus ja laadukkuus on ennakkoehto kaikelle mielekkäälle päätöksenteolle.

julkaistaan vahingossa. Nykyään valokuvia tai videoita saatetaan käyttää deepfake-videoiden tai muun disinformaation generointiin, jolloin tietojen vuotamisella voi olla yllättäviäkin kielteisiä seurauksia. Tietomurroilta ja -vuodoilta pyritään suojautumaan myös riittävällä tietoturvan käytännöllä.

Eettisyyttä tulisi tarkastella myös **läpinäkyvyyden ja osallisuuden näkökulmasta** pohtimalla, ymmärtävätkö palveluiden vaikutuspiirissä olevat henkilöt ja yhteisöt, miten palvelut toimivat, ja miten heitä koskevia tietoja käsitellään. Hyvin pitkiä ja juridista kieltä sisältäviä tietosuojaselosteita on syytetty siitä, että yksilöillä ei ole käytännössä mahdollisuutta ymmärtää datan käsittelyn periaatteita ennen palveluiden käyttöönottoa.

Kertaluontoisen suostumuksen antamisen sijaan hyvänä lähtökohtana voi pitää käytäntöä, jossa yksilö voi hallita itse lupia datansa käyttöön esimerkiksi palvelun käyttöliittymän kautta, ja tarvittaessa jopa siirtää datansa toiseen palveluun.

Lähdedatan vinoumat periytyvät herkästi tehtyihin johtopäätöksiin.

Ymmärtävätkö palveluiden vaikutuspiirissä olevat henkilöt ja yhteisöt, miten palvelut toimivat, ja miten heitä koskevia tietoja käsitellään?

Datastrategia ja tiedonhallinta mahdollisuuksina kädenjäljen lisäämiseen

Vastuullisuustyö on paitsi oman jalanjäljen minimoimista, myös oman kädenjäljen maksimointia. Vaikka edellä käsiteltyjen dataan liittyvien riskien tunnistaminen ja välttäminen ovat tärkeitä etiikasta keskusteltaessa, on hyvä muistaa myös positiivinen vaikuttavuus.

Organisaation kannattaa siksi vakavasti kysyä itseltään, minkälaista hyvää se voi datallaan tehdä. Taloudellisen lisäarvon tuottamisen lisäksi tavoitteita tulee pohtia myös asiakkaiden ja käyttäjien kokemuksen lisäarvon näkökulmasta. Esimerkiksi yritysten tarpeista ja työnhakijoiden osaamisesta dataa keräävä verkkopalvelu voi mahdollistaa kummankin osapuolen kannalta parempia rekrytointipäätöksiä, ja lisäksi siten myös yhteiskunnallista hyvinvointia.

Joskus organisaation haasteena voi olla juuri se, että mahdollisuuksia omaan kädenjälkeen ei ole onnistuttu kirkastamaan itselle esimerkiksi osana datastrategiaa. Dataa hyödynnettäessä kohdataan erilaisia riskejä, mutta jos dataa ei hyödynnetä lainkaan, jäävät digitaalisen teknologian positiiviset mahdollisuudet kokonaan hyödyntämättä.

Kun strategiset tavoitteet ovat olemassa, on luvassa paljon niin sanottua tiedon johtamisen työtä, jotta näihin tavoitteisiin päästään. Monet datan hyödyntämisen haasteet palautuvat tavalla tai toisella puutteisiin tiedon hallinnassa. Tiedonhallinnan suunnitelmassa tulisi dokumentoida ainakin se, mitä data sisältää, miksi se on kerätty, kuka siitä on vastuussa, kenelle data on or-

ganisaatiossa suunnattu, miten dataa päivitetään, ja milloin se (mahdollisesti) poistetaan. Samalla voidaan pohtia datan suotavia ja kiellettyjä käytötapoja, jotka auttavat myös uusia työntekijöitä ymmärtämään dataa ja sen tavoitteita.

Tiedonhallinnasta on puhuttu jo pitkään, mutta viime aikoina tekoälytyökalujen käyttöön-otto on muistuttanut monia sen merkityksestä. Jos organisaation dokumentteja aineistonaan hyödyntävä generatiivinen tekoäly tuottaa huonoja lopputuloksia, voi syy olla siinä, että dokumentteja ei ole luokiteltu yhteisten käytäntöjen mukaisesti. Tämän taustalla voi taas olla se, että datalle ei ole määritelty omistajia tai oikeaa yleisöä, jolla saisi olla pääsy dataan. Jos tekoäly ei löydä sopivaa aineistoa päätöstensä tueksi, ei sitä välttämättä löydä yksittäinen työntekijäkään.

Datavastuullisuus on sekä riskien tunnistamista että positiivisen vaikuttavuuden edistämistä. Vastuullisuus ei ole yksittäisen työntekijän harjoitus vaan yhdessä tekemistä. Keskustelut koskevat ainakin johtoa, tuotekehitystä, tietohallintoa, tiedon käsitteilyjä sekä tietosuoja- ja tietoturva-asiiantuntijoita. Jos eri rooleissa olevat henkilöt löytävät datasta keskusteluun yhteisen kielen ja määrittelevät yhdessä tärkeimmät kysymykset, otetaan iso askel kohti vastuunjakoja ja tehtävälisiä. ■

Vastuullisuus ei ole yksittäisen työntekijän harjoitus vaan yhdessä tekemistä.

Lisätietoa datavastuullisuuden valmennuksesta: <https://tieke.fi/palvelut/liiketoimintapalvelut/datavastuullisuuden-valmennus/>

Organisaation kannattaa vakavasti kysyä itseltään, minkälaista hyvää se voi datallaan tehdä.



Mikko Eloholma on taustaltaan kasvatustieteen maisteri (yleinen ja aikuiskasvatus), ja on työskennellyt TIEKellä vuodesta 2020 digiosaamisen vauhdittajana keskittyen erityisesti organisaatioiden ja yksilöiden osaamiseen tietosuojasta, dataliiketoiminnasta ja datavastuullisuudesta. Mikko suoritti 2022 Open Data Instituten Data Ethics Facilitator -valmennuksen, ja on ollut päävastuussa neljän TIEKEN datavastuullisuuden valmennuksen suunnittelusta ja toteutuksesta.

LÄHTEITÄ JA LUKUVINKKEJÄ

Open Data Institute (2021). What is the Data Ethics Canvas? Verkkosivu.

Viitattu 6.9.2024. <https://theodi.org/insights/tools/the-data-ethics-canvas-2021/>

Open Data Institute (2022). Data Ethics Maturity Model: benchmarking your approach to data ethics. Verkkosivu ja julkaisu.

Viitattu 6.9.2024. <https://theodi.org/insights/tools/data-ethics-maturity-model-benchmarking-your-approach-to-data-ethics/>

Open Data Institute (2023). Responsible data stewardship. Julkaisu. Viitattu 6.9.2024.

<https://theodi.org/insights/reports/defining-responsible-data-stewardship/>

Sitra (2019). Oma päätösvalta datan käytössä on ihmisille tärkeää. Tutkimuksen tulokset. Viitattu 6.9.2024.

<https://www.sitra.fi/artikkelit/oma-paatosvalta-datan-kaytossa-ihmisille-tarkeaa/>

Sitra (2020). Yritysvastuu ulottuu dataan – Näkökulmia ja ehdotuksia vastuullisen datan käytön edistämiseksi. Julkaisu.

<https://www.sitra.fi/julkaisut/yritysvastuu-ulottuu-dataan/>

TIEKE Tietoyhteiskunnan kehittämiskeskus ry. Datavastuullisuuden valmennus – hanki valmiudet vastuulliseen dataan ja tekoälyn hyödyntämiseen. Verkkosivu. Viitattu 6.9.2024. <https://tieke.fi/palvelut/liiketoimintapalvelut/datavastuullisuuden-valmennus/>