

Algoritmin perusteet ja tekoälyn etiikka -kurssi: Teoriasta käytäntöön

Työkalupakki synkronisille luennoille

Osaamiskokonaisuus 1 | Tekoälyn
etiikka
PowerPoint-diat tueksi

SISÄLLYS

- JOHDANTO - 3
- TEKOÄLYN EETTISET PERIAATTEET - 8
- TEKOÄLYN EETTISET VIITEKEHYKSET, OHJEET JA TYÖKALUPAKIT - 23
- EU:N TEKOÄLYLAKI (EU AI ACT) – 33
- TEKOÄLYN KEHITTÄMISEN ELINKAARI – 40
- SIDOSRYHMIEN OSALLISTUMINEN TEKOÄLYRATKAISUJEN KEHITTÄMISESSÄ – 45
- EETTISET PERIAATTEET JA SIDOSRYHMIEN OSALLISTUMINEN TEKOÄLYN KEHITTÄMISEN ELINKAARESSA - 53

JOHDANTO

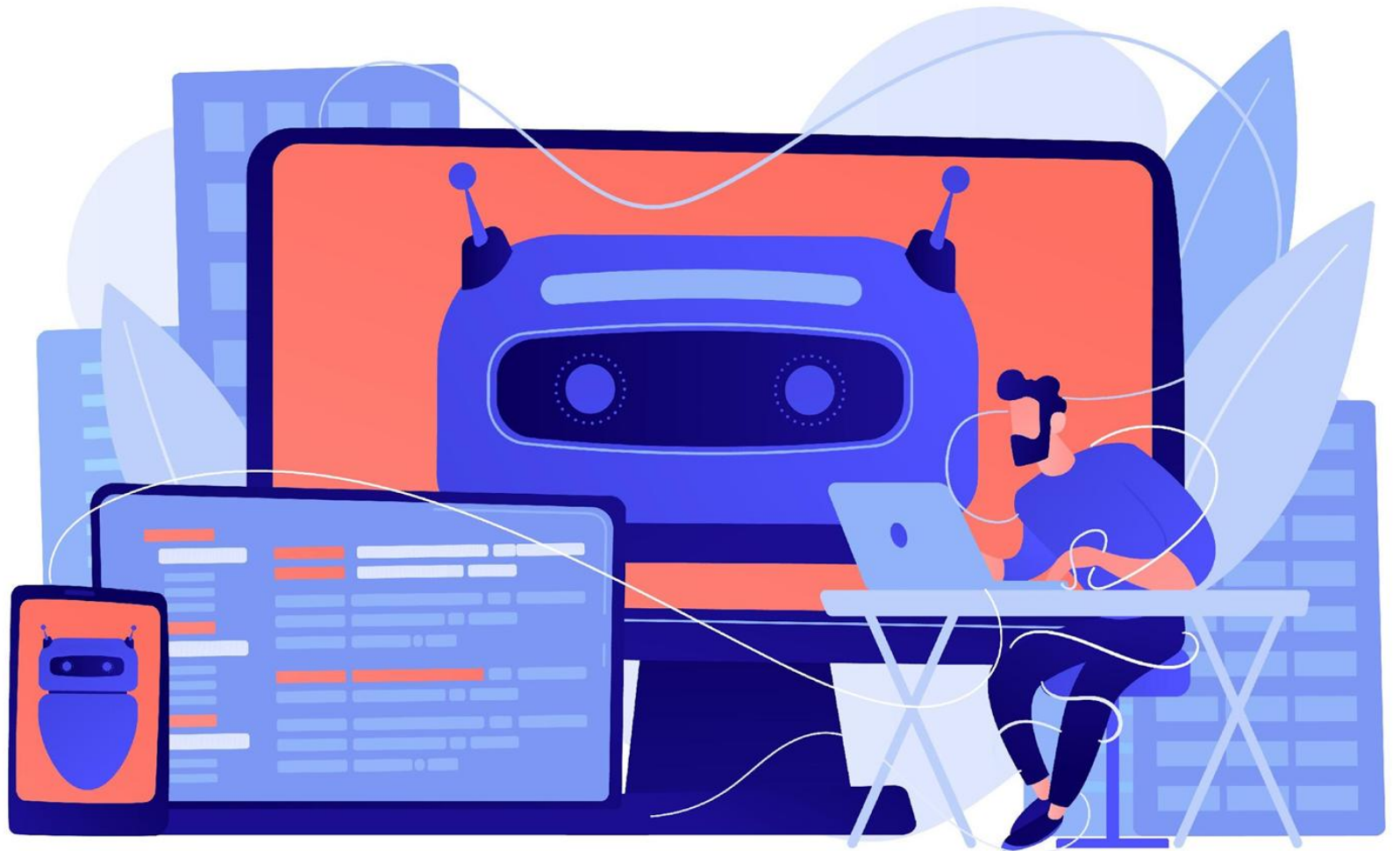


IMAGE SOURCE | Freepik

TÄSTÄ OSAAMISKOKONAISUUDESTA LÖYDÄTSEURAAVAT AIHEET:

- Tekoälyn etiikka ja sen merkitys tekoälyratkaisujen kehittämisessä.
- Tekoälyn eettiset periaatteet.
- Eettiset viitekehykset, ohjeistukset ja työkalupaketit.
- Korkean tason asiantuntijaryhmä; Luotettavan tekoälyn viitekehys.
- EU:n tekoälyasetus (EU AI Act).
- Tekoälyn kehityksen elinkaari.
- Sidosryhmien osallistaminen tekoälyn kehityksen elinkaaren aikana.
- Tekoälyn kehityksen elinkaari yhdistettynä eettisiin periaatteisiin ja sidosryhmiin.

OSAAMISKOKONAISUUDEN LOPUSSA SINUN TULISI OSATA:

- Tunnistaa tekoälyn kehittämiseen ja käyttöönottoon liittyvät eettiset periaatteet.
- Tunnistaa erilaisia eettisiä viitekehyksiä ja ohjeistuksia sekä ymmärtää, miten niitä sovelletaan tekoälyjärjestelmiin liittyvissä tosielämän tilanteissa.
- Ymmärtää eettisten periaatteiden ja sidosryhmien osallistamisen merkitys eettisessä tekoälyn kehitysprosessissa.

Eettiset huolet – Mitä tapahtuu, jos...

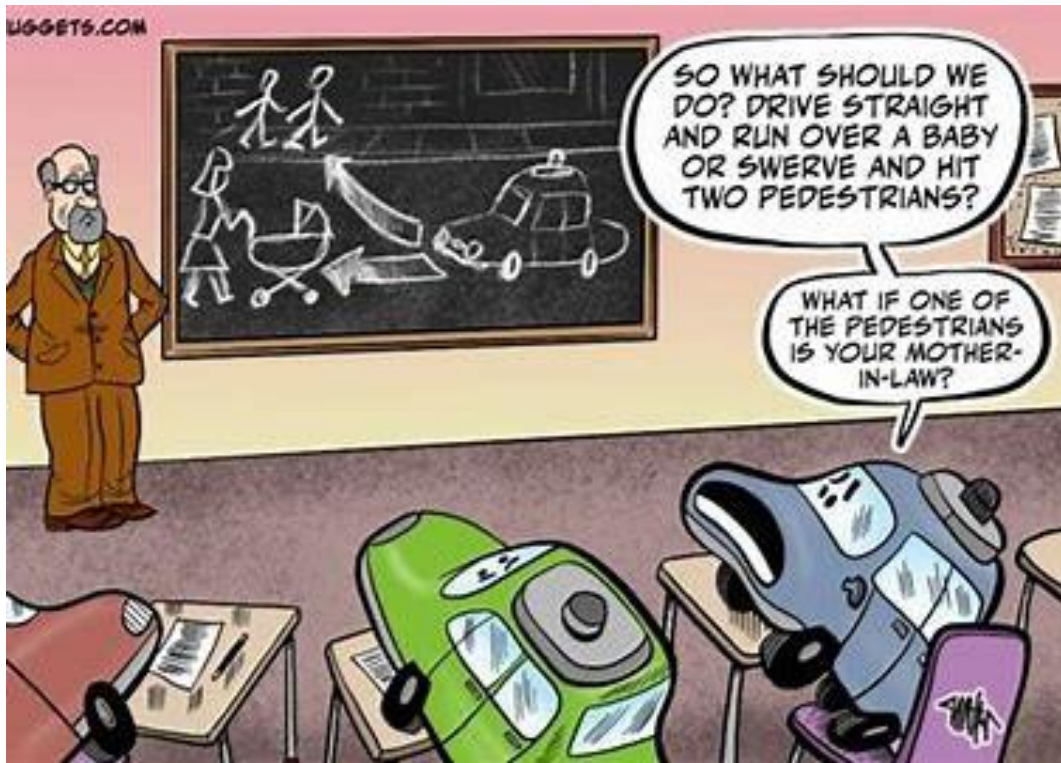


IMAGE SOURCE | [Cartoon: Teaching Ethics to AI - KDnuggets](#)

- Tekoälyllä olisi valta päättää, saatko lainaa?
- Tekoäly arvioisi, kuka tuomitaan vankeuteen?
- Autonominen auto joutuisi päättämään, pelastaako jalankulkijan hengen vai autossa olevien ihmisten hengen?
- Tekoälyyn perustuva rekrytointijärjestelmä päättäisi, palkataanko naishakija vai mieshakija lentokapteeniksi?

Tekoäly mullistaa elämää



IMAGE SOURCE | OECD; AI: Risk and Opportunity

Video from: <https://youtu.be/-CXkHs3cxa4>

- Tekoäly vaikuttaa kaikkiin elämän osa-alueisiin, kuten työhön ja vapaa-aikaan, ja tarjoaa ratkaisuja globaaleihin ongelmiin, kuten ilmastonmuutokseen ja terveydenhuollon saatavuuteen.
- Tekoälyn integroiminen talouteen ja yhteiskuntaan herättää kysymyksiä siitä, millaiset poliittiset ja institutionaaliset puitteet ovat tarpeen tekoälyn kehittämisen ja yhteiskunnan hyödyksi soveltamisen ohjaamiseksi.
- Laajat eettiset pohdinnat ovat johtaneet lukuisten lähestymistapojen luomiseen ja julkaisemiseen tekoälyn eettisen soveltamisen varmistamiseksi.

TEKOÄLYN EETTISET PERIAATTEET



Tekoälyn eettiset periaatteet

Mitä tekoälyn eettiset periaatteet ovat?

- **Eettiset periaatteet määritellään yleisinä ohjeina, jotka koskevat tekoälyjärjestelmien eettisiä kysymyksiä.** Nämä periaatteet ovat usein abstrakteja eivätkä anna tarkkoja ohjeita niiden soveltamisesta tietyssä tilanteessa.
- Eettiset periaatteet kuvaavat, **mitä oikeana ja vääränä pidetään** ja mitä muita eettisiä normeja noudatetaan. Tekoälyn eettiset periaatteet viittaavat normatiivisiin rajoituksiin, jotka koskevat algoritmien käytön sääntöjä yhteiskunnassa.

(Prem 2023), (Zhou et al., 2020)

Tekoälyn eettiset periaatteet

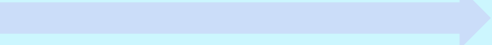
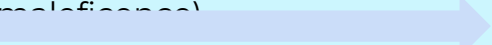
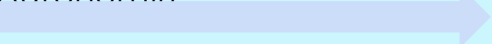
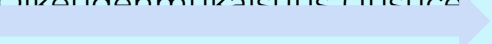
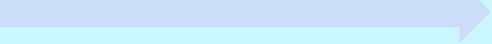
- Tekoälyn eettistä käyttöä ohjaavia suosituksia on julkaistu lukuisia.
- Suositukset ovat peräisin eri aloilta, esimerkiksi:
 - OECD:n suositus, tekoälyn eettisten periaatteiden pääpiirteet (julkaistu toukokuussa 2019)
 - Pekingin tekoälyn periaatteet tutkimukselle ja kehittämiselle (julkaistu toukokuussa 2019)
 - Teollisuus (esim. Google, IBM, Microsoft, Intel)
 - Hallitus (esim. Montrealin julistus, Euroopan komission asiantuntijaryhmä)
 - Akateeminen maailma (esim. Future of Life Institute, IEEE, AI4People)

(Prem 2023), (Morley et al 2020)

Tekoälyn eettiset periaatteet

Esimerkkejä eettisistä periaatteista

Kaavio osoittaa, miten eri viitekehykset yhdistävät eettiset periaatteet ja miten ne ovat päällekkäisiä. Vaakasuorat nuolet yhdistävät aiheet toisiinsa.

AI4People (julkaistu marraskuussa 2018) (Floridi ym. 2018)	Viisi periaatetta, jotka ovat keskeisiä tekoälyn eettisessä viitekehyksessä (L Floridi ja Clement-Jones 2019)	Luotettavan tekoälyn eettiset suuntaviivat (julkaistu huhtikuussa 2019) (Euroopan komissio 2019)	Tekoälyä koskeva suositus (julkaistu toukokuussa 2019) (OECD 2019b)
Hyvän tekeminen (beneficence) 	Tekoälyn on oltava hyödyllistä ihmiskunnalle	Ihmisen itsemääräämisoikeuden kunnioittaminen	Osallistava kasvu, kestävä kehitys ja hyvinvointi
Haitan välttäminen (non-maleficence) 	Tekoäly ei saa loukata yksityisyyttä tai heikentää turvallisuutta.	Haittojen ehkäiseminen	Kestävyys, turvallisuus ja varmuus
Autonomia 	Tekoälyn on suojeltava ja parannettava autonomiaamme ja kykyämme tehdä päätöksiä ja valita vaihtoehtojen välillä.		Ihmiskeskeiset arvot
Oikeudenmukaisuus (Justice) 	Tekoälyn on edistettävä vaurautta.	Oikeudenmukaisuus	Oikeudenmukaisuus
Selitettävyyt (Explicability) 	Tekoälyjärjestelmien on oltava ymmärrettäviä ja selitettäviä	Selitettävyyt	Läpinäkyvyys ja selitettävyyt Vastuullisuus

Modified from Morley et al 2020



IMAGE SOURCE | Image developed by VAMK

Eettiset tekoälyn periaatteet ovat suurelta osin samankaltaisia, mikä viittaa yhtenäisten eettisten suuntaviivojen syntymiseen.

- Hyödyllinen ja ihmisiä ja ympäristöä kunnioittava (hyvän tekeminen).
- Vankka ja turvallinen (haitan välttäminen).
- Ihmisarvoa kunnioittava (autonomia).
- Oikeudenmukainen (oikeudenmukaisuus).
- Selitettävä, vastuullinen ja ymmärrettävä (selitettävyyys).

Morley et al (2020), Jobin et al (2019)

Tekoälyn eettiset periaatteet



IMAGE SOURCE | Freepik

Hyvän tekeminen

Hyvän tekeminen tekoälyssä käsittää ihmiskunnan hyvinvoinnin edistämisen

Hyvän tekemisen periaate korostaa ihmisille ja planeetalle hyödyllisen tekoälyn luomista ja painottaa hyvinvoinnin ja ympäristön kestävyden edistämisen tärkeyttä.

(Floridi et al 2018)

Tekoälyn eettiset periaatteet



IMAGE SOURCE | Freepik

Esimerkki hyvän tekemisen periaatteesta

Terveydenhuollon suositukset

Ajattele tekoälypohjaista terveydenhuoltojärjestelmää, joka suosittelee hoito-ohjelmia potilastietojen perusteella. Jos tekoäly asettaa etusijalle hoitovaihtoehtot, jotka ovat terveydenhuollon tarjoajalle kannattavampia kuin potilaan terveydelle hyödyllisimmät vaihtoehtot, se ei noudata hyvän tekemisen periaatetta. Tässä tilanteessa tekoälyn suositukset priorisoivat taloudelliset hyödyt potilaan hyvinvoinnin maksimoinnin sijaan.

- IBM:n Watson for Oncology -järjestelmä kohtasi eettisiä haasteita, koska se suositteli hoitoja, jotka eivät aina olleet lääketieteellisten ohjeiden tai potilaan etujen mukaisia, mikä herätti huolta siitä, että voitto asetettiin potilaan tulosten edelle. Lue lisää [tästä linkistä](#).

(Floridi et al 2018)

Tekoälyn eettiset periaatteet



IMAGE SOURCE | Freepik

Haitan välttäminen

Haitan välttäminen tekoälyssä keskittyy haittojen välttämiseen.

Haitan välttäminen keskittyy haittojen ehkäisemiseen, olivatpa ne tahallisia ihmisen toimia tai tahattomia koneiden käyttäytymismalleja, mukaan lukien tahaton vaikutus ihmisen käyttäytymiseen. Kyse on negatiivisten tulosten välttämisestä riippumatta niiden lähteestä.

(Floridi et al 2018)

Tekoälyn eettiset periaatteet



IMAGE SOURCE | Freepik

Esimerkki haitan välttämisen periaatteesta

Ennakoiva poliisitoiminta

- Esimerkki tekoälyn autonomian heikentämisestä voisi olla sosiaalisen median suositusalgoritmi, joka hyödyntää personoitua sisältöä manipuloidakseen käyttäjien käyttäytymistä ja mielipiteitä ilman heidän nimenomaista suostumustaan. Syöttämällä käyttäjille jatkuvasti sisältöä, joka vahvistaa heidän olemassa olevia uskomuksiaan tai vinoumiaan, tekoäly rajoittaa heidän altistumistaan erilaisille näkökulmille ja vaikuttimille, heikentäen näin heidän autonomiaansa tehdä tietoisia päätöksiä laaja-alaisen tiedon pohjalta.
- Todellinen esimerkki: Facebookin uutisvirran algoritmi, joka on suunniteltu personoimaan sisältöä käyttäjien vuorovaikutuksen perusteella, on saanut kritiikkiä kuplien ja kaikukammioiden luomisesta. Tämän on katsottu mahdollisesti voimistavan sensaatiomaista tai polarisoivaa sisältöä ja vaikuttavan käyttäjien käyttäytymiseen ilman heidän suoraa suostumustaan. Lue lisää [tästä linkistä](#).

Tekoälyn eettiset periaatteet

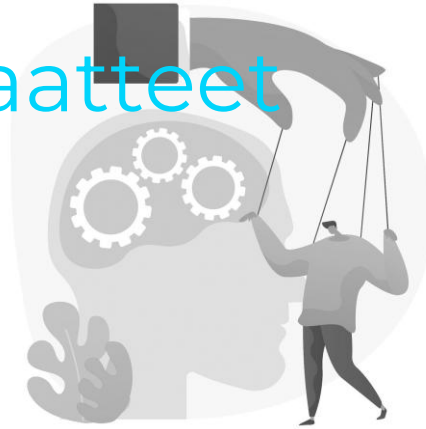


IMAGE SOURCE | Freepik

Autonomia

Tekoälyn autonomia edellyttää huolellista tasapainoa ihmisen ja koneen päätöksenteon välillä, korostaen tarvetta säilyttää ihmisen itsemääräämisoikeus.

- Autonomia edellyttää päätöksentekovallan tasapainottamista ihmisten ja tekoälyn välillä. Se puolustaa ihmisen valinnan sisäistä arvoa merkittävissä päätöksissä ja kannattaa "päättää delegoida" -mallia, jossa ihmisillä on viime kädessä valta delegoida ja tarvittaessa kumota tekoälyn päätökset.

(Floridi et al 2018)

Tekoälyn eettiset periaatteet



IMAGE SOURCE | Freepik

Esimerkki autonomian periaatteesta

Sosiaalisen median manipulointi

- Esimerkki tekoälyn autonomian heikentämisestä voisi olla sosiaalisen median suositusalgoritmi, joka hyödyntää personoitua sisältöä manipuloidakseen käyttäjien käyttäytymistä ja mielipiteitä ilman heidän nimenomaista suostumustaan. Syöttämällä käyttäjille jatkuvasti sisältöä, joka vahvistaa heidän olemassa olevia uskomuksiaan tai vinoumiaan, tekoäly rajoittaa heidän altistumistaan erilaisille näkökulmille ja vaikuttimille, heikentäen näin heidän autonomiaansa tehdä tietoisia päätöksiä laaja-alaisen tiedon pohjalta.
- Todellinen esimerkki: Facebookin uutisvirran algoritmi, joka on suunniteltu personoimaan sisältöä käyttäjien vuorovaikutuksen perusteella, on saanut kritiikkiä kuplien ja kaikkukammioiden luomisesta. Tämän on katsottu mahdollisesti voimistavan sensaatiomaista tai polarisoivaa sisältöä ja vaikuttavan käyttäjien käyttäytymiseen ilman heidän suoraa suostumustaan. Lue lisää [tästä linkistä](#).

Bias in algorithms – Artificial intelligence and discrimination (europa.eu) / Electronic Frontier Foundation (EFF), Pew Research Center, or Data & Society CUT | AI ethics - a practical approach

Tekoälyn eettiset periaatteet



IMAGE SOURCE | Freepik

Oikeudenmukaisuus

Tekoälyn oikeudenmukaisuuden periaate koskee päätöksenteon tasapuolisuutta ja sen jakautumista yhteiskunnassa, ja sen tavoitteena on puuttua eriarvoisuuteen ja edistää oikeudenmukaisuutta.

- Oikeudenmukaisuuden periaate koskee tekoälyn käyttöä historiallisten epäoikeudenmukaisuuksien korjaamisessa, tekoälyn hyötyjen tasapuolisen jakautumisen varmistamisessa sekä uusien haittojen tai sosiaalisen yhteenkuuluvuuden häiriöiden ehkäisemisessä.

(Floridi et al 2018)

Tekoälyn eettiset periaatteet



IMAGE SOURCE | Freepik

Esimerkki oikeudenmukaisuuden periaatteesta

Rikostuomioiden tekoälyalgoritmit

- Rikostuomioihin käytetyt tekoälypohjaiset algoritmit voivat tahattomasti ylläpitää vinoumia tiettyjä väestöryhmiä kohtaan. Jos algoritmien kouluttamiseen käytetty historiallinen data heijastaa ihmisten tekemiä puolueellisia päätöksiä, järjestelmä saattaa suositella pidempiä tuomioita tai ankarampia rangaistuksia syrjäytyneisiin yhteisöihin kuuluville henkilöille, pahentaen näin rikosoikeusjärjestelmän olemassa olevia eriarvoisuuksia.
- COMPAS-järjestelmän, jota on käytetty rikostuomioiden riskiarvioinnissa, havaittiin osoittavan rodullista vinoumaa: se luokitteli tummaihoiset syytetyt suhteettoman usein korkeamman riskin omaaviksi, mikä vaikutti epäoikeudenmukaisiin tuomioihin. Lue lisää [tästä linkistä](#).

Megan T. Stevenson and Christopher Slobogin, University of Pennsylvania Law Review

Tekoälyn eettiset periaatteet



IMAGE SOURCE | Freepik

Selitettävyyys

Selitettävyyys tekoälyssä tarkoittaa sitä, että ymmärrämme, miten tekoäly tekee päätöksiä, ja pystymme selittämään nämä päätökset selkeästi. Se varmistaa, että tekoälyjärjestelmät ovat läpinäkyviä ja vastuussa toimistaan.

(Floridi et al 2018)

Tekoälyn eettiset periaatteet



IMAGE SOURCE | Freepik

Esimerkki selitettävyyden periaatteesta

Luottopisteytysalgoritmit

- Kuvitellaan tekoälyllä toimiva luottopisteytysjärjestelmä, joka arvioi yksilöiden luottokelpoisuutta, mutta jonka päätöksentekoprosessi ei ole läpinäkyvä. Jos hakijoilta evätään lainoja tai heille tarjotaan epäedullisia ehtoja ilman ymmärrystä siitä, mihin tekijöihin tekoälyalgoritmin arvio perustuu, se heikentää heidän mahdollisuuksiaan kyseenalaistaa tai valittaa päätöksistä. Ilman selkeitä selityksiä siitä, miten tekoäly arvioi luottoriskiä, yksilöt saattavat kokea tulleeensa kohdelluiksi epäoikeudenmukaisesti ja menettää luottamuksensa järjestelmän reiluuteen.
- Equifaxin tietovuoto vuonna 2017 paljasti miljoonien kuluttajien arkaluonteisia tietoja ja nosti esiin huolia luottopisteytysalgoritmien läpinäkyvyydestä ja tietoturvasta. Tapaus korosti laajempia alan ongelmia, jotka liittyvät kuluttajien rajalliseen näkyvyyteen heidän luottokelpoisuuteensa vaikuttavista tekijöistä. Lue lisää [tästä linkistä.](#)

Keith Frankish & William M. Ramsey

EETTISET VIITEKEHYKSET, OHJEET JA TYÖKALUPAKIT



Eettiset viitekehykset, ohjeet ja työkalupakit

Eettiset viitekehykset

- Eettiset viitekehykset ovat enemmän kuin korkean tason periaatteita. Ne tarjoavat jäsennellyn lähestymistavan eettiseen päätöksentekoon.
- Monet tekoälyn eettiset viitekehykset pyrkivät tunnistamaan mahdollisia eettisiä haasteita ja ehdottamaan ratkaisuja näiden haasteiden voittamiseksi tai niihin liittyvien riskien lieventämiseksi.
- Tekoälyn eettinen viitekehys toimii perustana lainsäädännön, sääntöjen, teknisten standardien ja parhaiden käytäntöjen muotoilulle eri aloilla ja alueilla.

Eettisten viitekehysten tyypillisiä osia ovat:

- Käsitteet – merkityksellisiä eettisten näkökohtien keskustelussa
- Periaate – eettiset periaatteet (esim. arvot)
- Huolenaihe – miten periaatteet ovat uhattuina tekoälyjärjestelmien käytön ja kehittämisen kautta
- Korjaava toimenpide – strategiat, säännöt ja ohjeet huolenaiheiden ratkaisemiseksi

(Floridi & Cows, 2019), (Prem, 2023: 701) (Ayling & Chapman, 2022)

Eettiset viitekehykset, ohjeet ja työkalupakit

Eettiset ohjeet ja työkalupakit

- Eettiset ohjeet tarjoavat käytännön ohjeita ja erityisiä sääntöjä eettiseen käyttäytymiseen eri tilanteissa.
- Eettiset työkalupakit ja ohjeet muokkaavat tekoälyn perustuvaa innovointia ja tukevat tekoälyn eettisten periaatteiden käytännön soveltamista.
- Työkalupakit ja ohjeet selittävät, miten eettisiä periaatteita sovelletaan tekoälyn suunnittelussa, toteutuksessa ja käyttöönotossa.
- Viitekehykset, ohjeet ja työkalupakit ovat usein päällekkäisiä.
- Seuraavilla sivuilla esitellään tarkemmin kaksi esimerkkiä viitekehysistä: HLEG Trustworthy AI Framework ja OECD AI Classification Framework.

(Zhou et al., 2020)

Eettiset viitekehykset, ohjeet ja työkalupakit

Korkean tason asiantuntijaryhmän luotettavan tekoälyn viitekehys



- Vaikka tekoälyn eettisyydestä on päästy yksimielisyyteen, siitä, mitä "eettinen tekoäly" tarkoittaa ja mitä eettisiä vaatimuksia ja teknisiä standardeja sen saavuttamiseksi tarvitaan, käydään edelleen keskustelua.
- Euroopan komissio esitti strategian tekoälyn kehittämiseksi Euroopassa, jossa korostetaan eettistä, turvallista ja edistyksellistä tekoälyä.
- AI-HLEG esitti huhtikuussa 2019 luotettavaa tekoälyä koskevat eettiset suuntaviivat luotettavan tekoälyn edistämiseksi.
- Luotettava tekoäly on tunnustettu olennaiseksi tavoitteeksi luottamuksen lisäämiseksi tekoälyn kehittämiseen ja soveltamiseen. Se perustuu vankkaan kehykseen, joka varmistaa tekoälyn luotettavuuden.
- Lisätietoja <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>

(Leijnen et al., 2020), HLEG (2019). High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. Brussels: European Commission

Eettiset viitekehykset, ohjeet ja työkalupakit

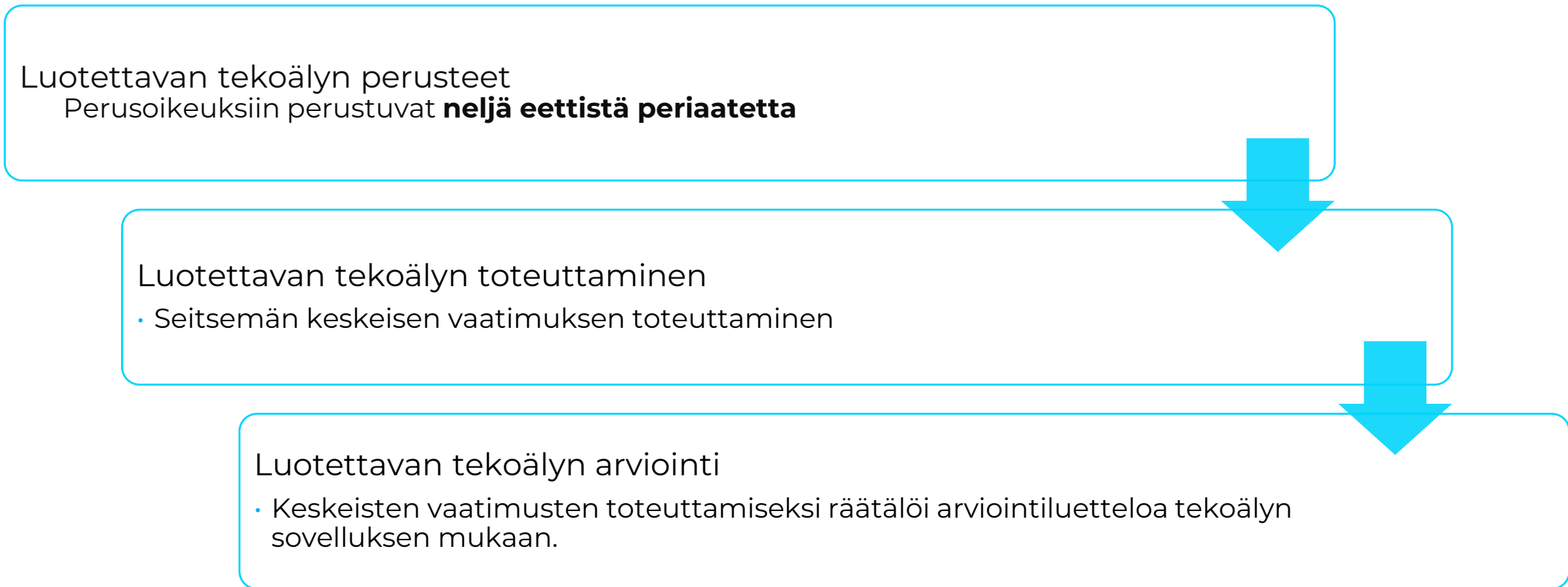
Luotettavan tekoälyn viitekehys

Luotettavalla tekoälyllä on kolme osa-aluetta, jotka tulee täyttyä koko järjestelmän elinkaaren ajan:

- Laillinen – noudattaa kaikkia sovellettavia lakeja ja säädöksiä
- Eettinen – kunnioittaa eettisiä periaatteita ja arvoja
- Kestävä – teknisesti luotettava ja ottaa huomioon myös sosiaalisen toimintaympäristönsä

Eettiset viitekehykset, ohjeet ja työkalupakit

Korkean tason asiantuntijaryhmän luotettavan tekoälyn viitekehys



HLEG (2019). High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. Brussels: European Commission.

Eettiset viitekehykset, ohjeet ja työkalupakit

Luotettavan tekoälyn perusteet: Eettiset periaatteet

Ihmisen autonomian kunnioittaminen

- Tekoälyn tulisi vahvistaa itsemääräämisoikeutta, välttää manipulointia ja tukea ihmisten kognitiivisia, sosiaalisia ja kulttuurisia taitoja. Sen on mahdollistettava ihmisten valvonta ja merkitykselliset valinnat sekä tuettava

Haitan välttäminen

- Tekoäly ei saa aiheuttaa haittaa, sen on suojeltava ihmisarvoa, oltava turvallista ja luotettavaa, ja siinä on otettava erityisen huomioon haavoittuvat ryhmät ja ympäristö. Sen olisi ehkäistävä valta- tai tietojärjestelmien

Oikeudenmukaisuus

- Tekoälyn tulisi varmistaa etujen ja kustannusten tasapuolinen jakautuminen, estää vinoumat ja syrjintä sekä edistää yhdenvertaisia mahdollisuuksia. Se edellyttää keinojen ja päämäärien tasapainoa ja mahdollistaa tekoälyn

Selitettävyyys

- Tekoälyprosessien on oltava läpinäkyviä, ja niiden kyvykkyydet ja tarkoitukset on ilmoitettava selkeästi. Kun täydellistä selitettävyyttä ei ole mahdollista saavuttaa, oikeuksien kunnioittaminen on varmistettava muilla toimenpiteillä, kuten jäljitettävyydellä ja auditoitavuudella.

- AI HLEG luettelee neljä perusoikeuksiin perustuvaa eettistä periaatetta, joita on noudatettava, jotta tekoälyjärjestelmät kehitetään, otetaan käyttöön ja käytetään luotettavalla tavalla.
- Eettiset periaatteet on määritelty eettisiksi velvoitteiksi, joita tekoälyn ammattilaisten on aina pyrittävä noudattamaan.
- AI HLEG tunnustaa ja käsittelee myös näiden periaatteiden välisiä mahdollisia jännitteitä.

<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Eettiset viitekehykset, ohjeet ja työkalupakit

Luotettavan tekoälyn toteuttaminen: keskeiset vaatimukset

- Vahvistetut eettiset periaatteet on muutettava lopulliseksi vaatimuksiksi luotettavan tekoälyn toteuttamiseksi. Nämä vaatimukset koskevat kaikkia tekoälyjärjestelmän elinkaaren osapuolia:
- Kehittäjät: tekoälyn tutkimusta, suunnittelua tai kehittämistä tekevät henkilöt tai tiimit.
- Käyttäjät: tekoälyä toiminnassaan hyödyntävät ja tekoälypohjaisia tuotteita tai palveluja tarjoavat organisaatiot.
- Loppukäyttäjät: tekoälyjärjestelmien kanssa suoraan tai välillisesti vuorovaikutuksessa olevat henkilöt.
- Yhteiskunta: tekoälyn vaikutusten kohteena olevat henkilöt, sekä suoraan että välillisesti.

Ihmisen toiminta ja valvonta

- perusoikeudet, ihmisen toimijuus ja ihmisen valvonta

Tekninen kestävyys ja turvallisuus

- hyökkäysten kestävyys ja turvallisuus, varasuunnitelma ja yleinen turvallisuus, tarkkuus, luotettavuus ja toistettavuus

Tietosuoja ja tietojen hallinta

- tietojen yksityisyyden, laadun ja eheyden kunnioittaminen sekä tietojen säilyttäminen

Läpinäkyvyys

- jäljitettävyys, selitettävyys ja viestintä

Monimuotoisuus, syrjimättömyys ja oikeudenmukaisuus

- epäoikeudenmukaisen vinouman välttäminen, esteettömyys ja universaali suunnittelu sekä sidosryhmien osallistuminen

Yhteiskunnan ja ympäristön hyvinvointi

- kestävyys ja ympäristöystävällisyys, yhteiskunnallinen vaikutus, yhteiskunta ja demokratia

Vastuullisuus

- tarkastettavuus, kielteisten vaikutusten minimointi ja raportointi, kompromissit ja korjaavat toimenpiteet.

HLEG (2019). High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. Brussels: European Commission

Eettiset viitekehykset, ohjeet ja työkalupakit

Luotettavan tekoälyn arviointi: ALTAI-arviointiluettelo

- Luotettavan tekoälyn kehittämisen tueksi on kehitetty työkalu, joka muuntaa nämä tekoälyn periaatteet käytännön toimenpiteiksi. Tämä työkalu, joka tunnetaan nimellä Assessment List for Trustworthy AI (ALTAI), tarjoaa kattavan ja dynaamisen tarkistuslistan, joka toimii kehittäjille ja käyttöönottajille tiekarttana näiden periaatteiden integroimiseksi tekoälysovelluksiinsa.
- ALTAI helpottaa tätä prosessia tarjoamalla joukon konkreettisia vaiheita itsearviointia varten ja varmistaa siten, että tekoälyteknologiat kehitetään tavalla, joka maksimoi käyttäjien hyödyt ja minimoi tarpeettomien riskien altistumisen.
- [ALTAI-työkalu](#)



IMAGE SOURCE | ALTAI

(High-Level Expert Group on AI (AI HLEG), 2020)

Eettiset viitekehykset, ohjeet ja työkalupakit

OECD:n tekoälyn luokittelukehys

- OECD:n tekoälyn luokittelukehys, jonka on laatinut OECD:n tekoälyn asiantuntijaverkosto, on tarkoitettu auttamaan eri sidosryhmiä arvioimaan erilaisten tekoälyjärjestelmien potentiaalia ja riskejä.
- Tämä työkalu helpottaa tekoälystrategioiden kehittämistä ja pyrkii ylläpitämään politiikan johdonmukaisuutta kansainvälisellä tasolla.
- Kehys perustuu OECD:n tekoälyn periaatteisiin, joissa vahvistetaan oikeudenmukaisuuden, avoimuuden, turvallisuuden ja vastuullisuuden arvot sekä edistetään inhimillisten valmiuksien kehittämistä ja kansainvälistä yhteistyötä.



IMAGE SOURCE | OECD Framework

<https://www.oecd.org/digital/artificial-intelligence/>

EU:n tekoälylaki (EU AI ACT)



EU:n tekoälylaki (EU AI ACT)

- Tekoälylaki on tekoälyä koskeva EU:n lakiesitys. Tekoälylaki on ensimmäinen laatuaan maailmassa. Laki on tarkoitus saattaa voimaan vuonna 2026.
- Sitä sovelletaan tekoälyn kehittämiseen, käyttöönottoon ja käyttöön EU:ssa tai silloin, kun se vaikuttaa EU:n kansalaisiin.
- Euroopan komissio julkaisi ehdotuksen tekoälyä koskevaksi asetukseksi (tekoälyasetus) 21. huhtikuuta 2021. Asetus on ensimmäinen lainsäädäntökehys, jossa vahvistetaan yhdenmukaistetut säännöt tekoälyn kehittämiselle, markkinoille saattamiselle ja käytölle Euroopan unionissa.



IMAGE SOURCE | [The AI Act Explorer](https://theaiactexplorer.com/) | [EU Artificial Intelligence Act](https://www.euaiact.com)

<https://artificialintelligenceact.com>
<https://www.euaiact.com>

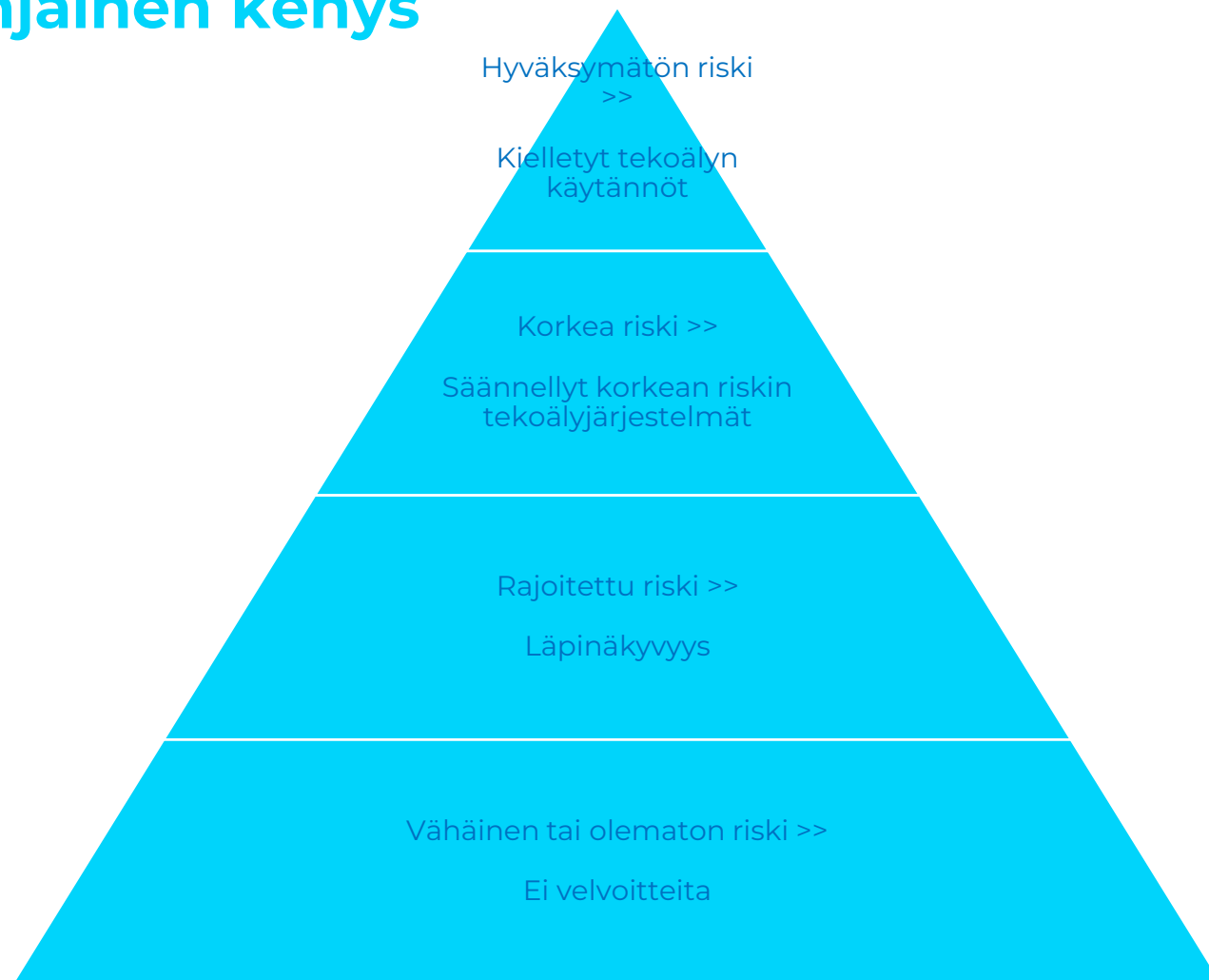
EU:n tekoäylaki (EU AI ACT)

Riskipohjainen kehys

Vahvistetut eettiset periaatteet on muutettava lopullisiksi vaatimuksiksi luotettavan tekoälyn toteuttamiseksi.

Nämä vaatimukset koskevat kaikkia tekoälyjärjestelmän elinkaaren osapuolia:

- Kehittäjät: tekoälyn tutkimusta, suunnittelua tai kehittämistä tekevät henkilöt tai tiimit.
- Käyttäjät: tekoälyä toiminnassaan hyödyntävät ja tekoälypohjaisia tuotteita tai palveluja tarjoavat organisaatiot.
- Loppukäyttäjät: tekoälyjärjestelmien kanssa suoraan tai välillisesti vuorovaikutuksessa olevat henkilöt.
- Yhteiskunta: tekoälyn vaikutusten kohteena olevat henkilöt, sekä suoraan että välillisesti.



The EU AI Act 'Pyramid of Risks' (Source: European Parliament)

EU:n tekoälylaki (EU AI ACT)

Riskipohjainen kehys | Vähäinen tai olematon riski



Tekoälysovellukset, joiden katsotaan aiheuttavan vain vähäisen tai ei lainkaan riskiä kansalaisten oikeuksille tai turvallisuudelle. Suurin osa tekoälyjärjestelmistä kuuluu **vähäisen riskin** luokkaan.

Esimerkkejä vähäisen riskin tekoälysovelluksista:

- Tekoälypohjaiset suosittelujärjestelmät
- Roskapostisuodattimet

Nykyisten säännösten mukaan näihin järjestelmiin ei sovelleta uusia erityisvelvoitteita, mutta niiden on oltava olemassa olevien lakien mukaisia.

EU:n tekoäylaki (EU AI ACT)

Riskipohjainen kehys | Rajoitettu riski



Rajoitetun riskin tekoälyjärjestelmät edellyttävät läpinäkyvyyttä.

Näiden tekoälyjärjestelmien vaatimukset koskevat läpinäkyvyyttä: Käyttäjille on ilmoitettava, että he ovat vuorovaikutuksessa tekoälyn kanssa, ja heillä on oltava käytettävissään tarvittavat tiedot käytön jatkamisesta päättämiseksi.

Esimerkkejä rajoitetun riskin tekoälysovelluksista:

- Tekoäly, joka tuottaa tai manipuloi kuva- tai äänisisältöä.
- Tekoäly, joka luo tai muuttaa videosisältöä, kuten deepfake-videoita.

EU:n tekoälylaki (EU AI ACT)

Riskipohjainen kehys | Korkea riski



Korkean riskin tekoälyjärjestelmät voivat vaikuttaa merkittävästi käyttäjän elämänmahdollisuuksiin.

Näitä järjestelmiä koskevat tiukat vaatimukset, jotka on täytettävä ennen niiden käyttöönottoa EU:n markkinoilla, mukaan lukien riskienhallintaa ja tietojen hallintaa koskevat velvoitteet.

Esimerkkejä korkean riskin tekoälystä ovat:

- kriittisen infrastruktuurin hallinta ja käyttö
- koulutus ja ammatillinen koulutus
- työllisyys, työntekijöiden hallinta ja pääsy itsenäiseen ammatinharjoittamiseen
- peruspalveluiden ja julkisten palveluiden ja etuuksien saatavuus ja käyttö
- lainvalvonta
- Muuttoliike, turvapaikka-asiat ja rajavalvonta
- Avustaminen lain tulkinnassa ja soveltamisessa.

EU:n tekoälylaki (EU AI ACT)

Riskipohjainen kehys | Hyväksymätön riski



Tekoälyjärjestelmät, joiden riskit ovat kohtuuttomat, katsotaan selkeäksi uhaksi ihmisten perusoikeuksille ja ne ovat kiellettyjä. Kohtuuttoman riskin järjestelmät ovat sellaisia, jotka ovat hyväksikäyttäviä, manipuloivia tai käyttävät alitajuisia tekniikoita.

Tällaisten tekoälyjärjestelmien myynti EU:n markkinoilla on kielletty.

Esimerkkejä kielletystä tekoälyn käytöstä ovat

- ihmisten tai tiettyjen haavoittuvien ryhmien kognitiivinen käyttäytymisen manipulointi: esimerkiksi ääniohjatut lelut, jotka rohkaisevat lapsia vaaralliseen käyttäytymiseen
- sosiaalinen pisteytys: ihmisten luokittelu käyttäytymisen, sosioekonomisen aseman tai henkilökohtaisten ominaisuuksien perusteella
- ihmisten biometrinen tunnistaminen ja luokittelu
- reaaliaikaiset ja etäbiometriset tunnistusjärjestelmät, kuten kasvojentunnistus

TEKOÄLYN KEHITTÄMISEN ELINKAARI



Tekoälyn kehittämisen elinkaari

Eettisten periaatteiden, viitekehysten ja suuntaviivojen muuttaminen konkreettisiksi toimiksi

Periaatteiden, puitteiden ja suuntaviivojen merkitys tekoälyn kehittämisessä on tunnustettu, mutta niiden soveltaminen käytännössä tekoälyratkaisujen luomisessa on monin tavoin haasteellista:

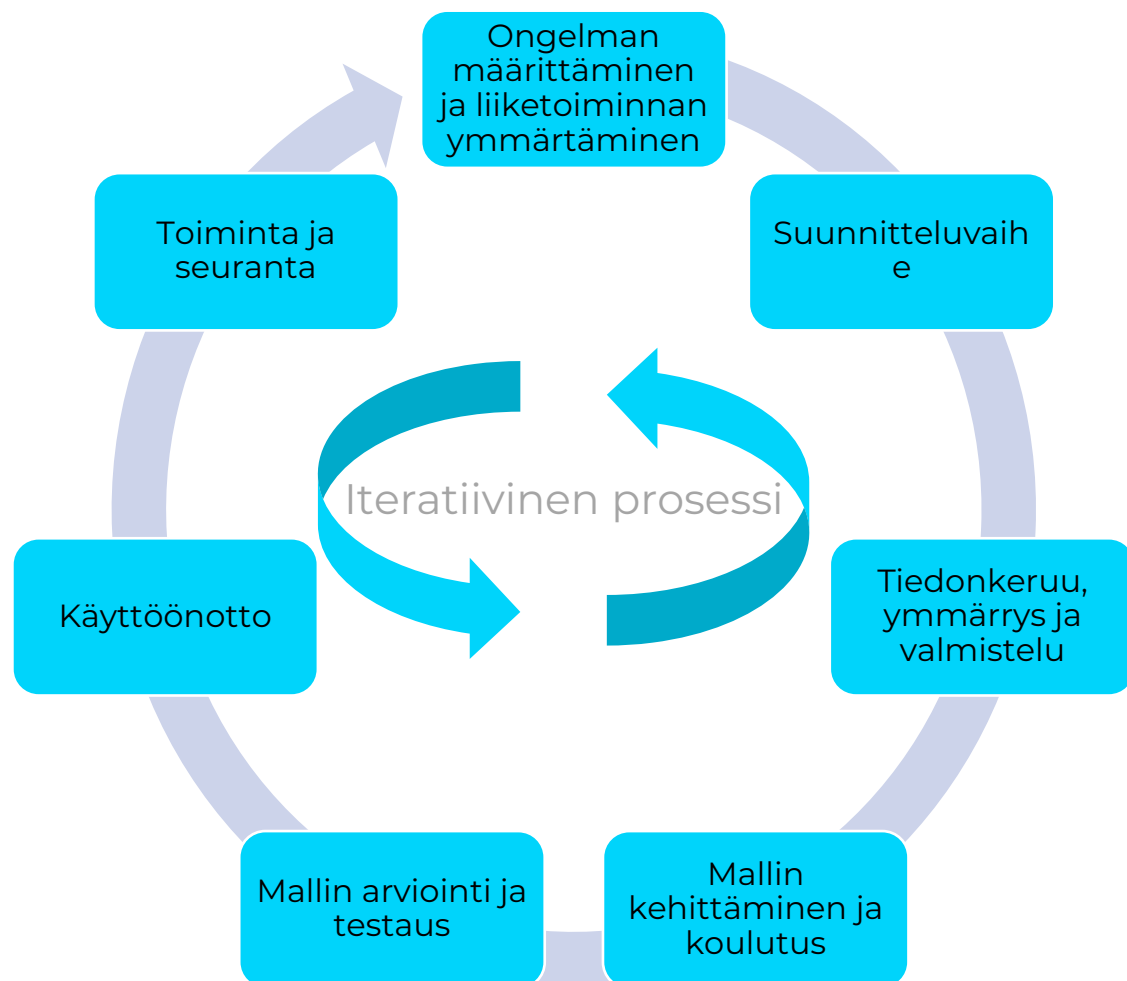
- Tekoälyn periaatteet ovat usein abstrakteja, moninaisia ja monimutkaisia.
- Ihmisen käyttäytymisen vivahteiden muuntaminen suoraviivaisiksi, yleispäteviksi työkaluiksi on monimutkaista.
- Periaatteet ovat usein henkilökohtaisen tulkinnan alaisia ja riippuvat yksilön taustasta.
- Säännöksissä, viitekehyksissä ja ohjeissa ei ole yleistä standardointia (vaikka joitakin yleistyksiä onkin tehty, kuten aiemmin esitettiin).
- Vaikka viitekehykset ovat hyödyllisiä ongelmien tunnistamisessa, ne eivät välttämättä tarjoa käytännön ohjeita konkreettisista toimista.

Tekoälyn kehittämisen elinkaari

Eettisten periaatteiden, viitekehysten ja suuntaviivojen muuttaminen konkreettisiksi toimiksi: Tekoälyn kehittämisen elinkaari

- Käytettävissä on tekoälyratkaisujen eettisiin haasteisiin vastaavia työkaluja, mutta ne kattavat usein vain valikoituja periaatteita tai tekoälyratkaisujen kehitysvaiheita.
- Eettisten periaatteiden muuttamiseksi konkreettisiksi toimiksi on suositeltavaa tarkastella tekoälyn kehitystä vaiheiden sarjana, joka alkaa liiketoimintatapauksesta ja päättyy ratkaisun toimintatapaan. (Tekoälyn kehityksen elinkaari)
- Tekoälyn kehityksen elinkaari ottaa huomioon eettiset näkökohdat kokonaisvaltaisesti kaikissa kehitysvaiheissa.

Tekoälyn kehittämisen elinkaari



- Kuten minkä tahansa projektin, myös tekoälyn suunnitteluprojektin on suositeltavaa noudattaa suunnitteluvaiheita, jotta innovatiiviset ideat voidaan muuttaa käytännöllisiksi, hyvin määritellyiksi ratkaisuiksi.
- On tärkeää ymmärtää, että jokainen vaihe on tärkeä osa ratkaisun kehittämistä ja muodostaa tekoälyn kehityksen elinkaaren.
- Prosessi on ITERATIIVINEN, mikä tarkoittaa, että tarvittaessa voidaan palata edelliseen vaiheeseen ja muokata ratkaisua.

Tässä esitetyt tekoälyn kehitysvaiheet ovat yhdistelmä eri malleja, jotka on kuvattu Data Science Process Alliance (2023), Morley et.al' (2019), Prehm's (2022) ja Rochel et.al'a (2020) artikkeleissa.

Data Science Process Alliance /Morley (2019) / Prem (2023) / Rochel & Evequoz (2020)

Tekoälyn kehittämisen elinkaari



Ongelman
määrittäminen ja
liiketoiminnan
ymmärtäminen

Ongelman määrittely on ratkaisevan tärkeää asiakkaan tarpeiden ja tekoälyratkaisun roolin ymmärtämiseksi.



Suunnitteluvaihe

Suunnitteluvaiheessa liiketoimintavaatimukset muutetaan insinöörien suunnitteluvaatimuksiksi.



Tiedonkeruu,
ymmärrys ja
valmistelu

Tietojen kerääminen, ymmärtäminen ja valmistelu ovat ratkaisun perusta.



Mallin
kehittäminen ja
koulutus

Mallin kehittäminen ja arviointi tähtäävät siihen, että tiedot voidaan muuttaa alkuperäisen ongelman ratkaisuksi.



Mallin arviointi ja
testaus

Malli on testattava perusteellisesti ja sen toiminnallisuus on sovitettava asiakkaan tarpeisiin.



Käyttöönotto

Käyttöönotto tuo ratkaisun käyttäjien saataville.



Toiminta ja
seuranta

Ratkaisua on seurattava ajan mittaan ja ylläpidettävä muuttuvien tarpeiden mukaisesti.

SIDOSRYHMIEN OSALLISTUMINEN TEKOÄLYRATKAISUJEN KEHITTÄMISEEN

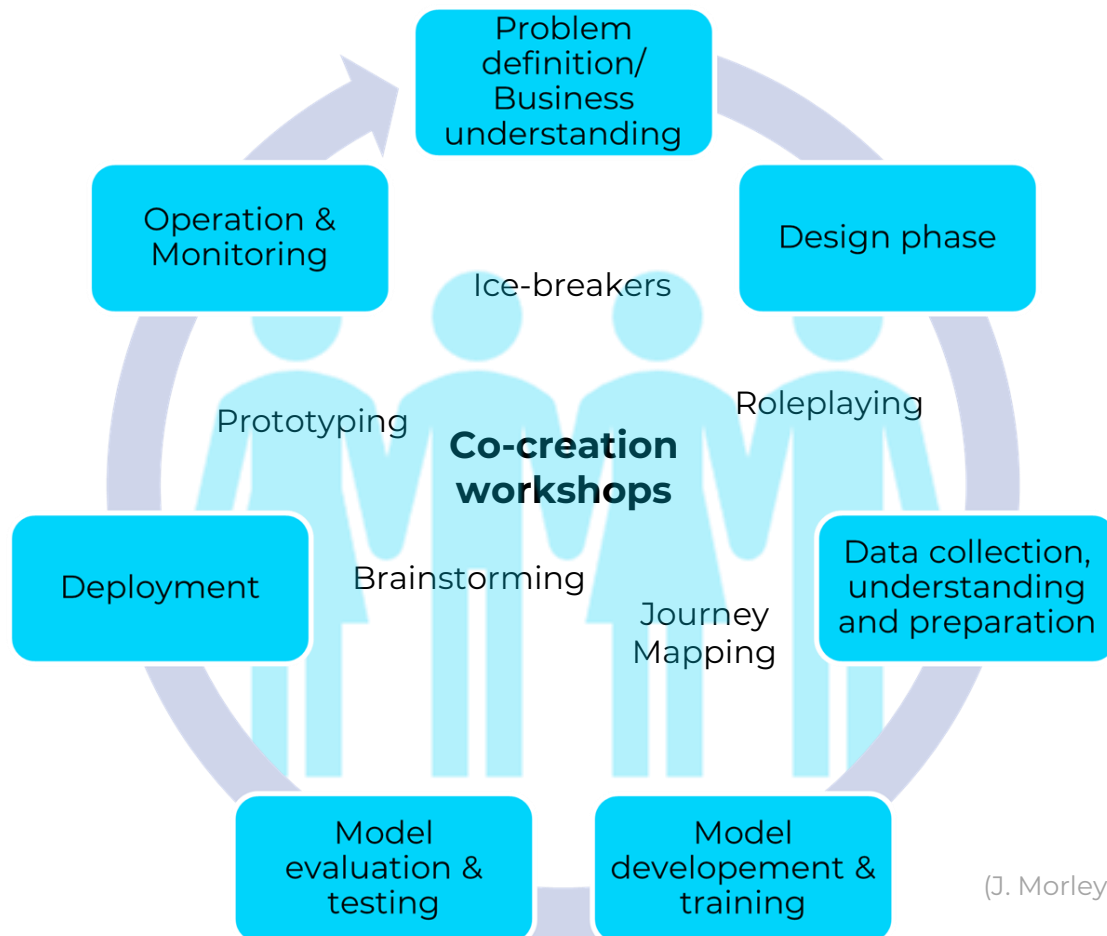


IMAGE SOURCE | Freepik

Sidosryhmien osallistuminen tekoälyratkaisujen

kehittämiseen

Miksi on tärkeää ottaa eri sidosryhmät mukaan tekoälyratkaisujen kehittämiseen?

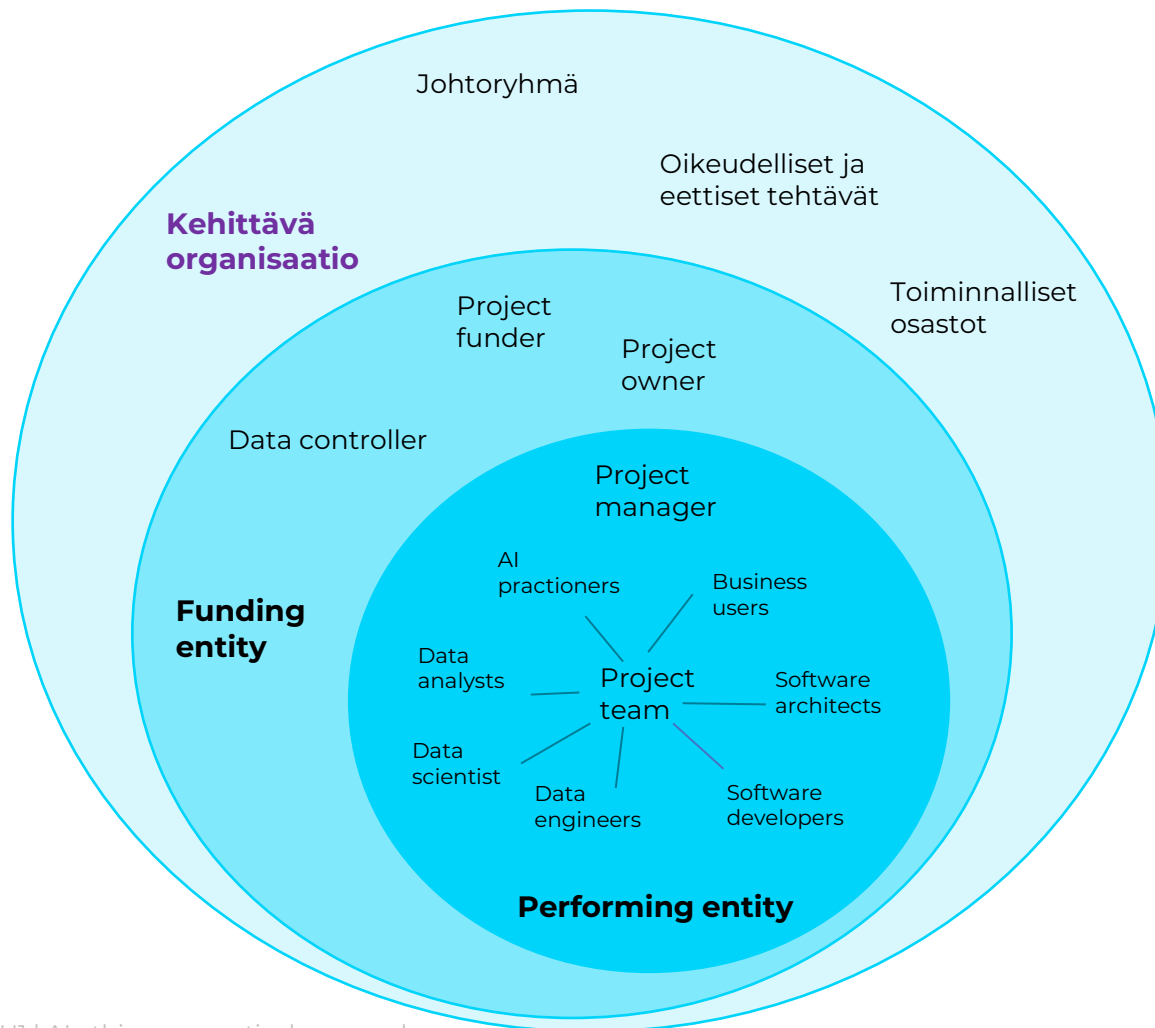


- Jotta tekoälyn kehittämisen vaikutukset voidaan ottaa huomioon eri näkökulmista ja kehittää luotettavia järjestelmiä, on välttämätöntä ottaa mukaan ja kuulla useita sidosryhmiä, mukaan lukien ne, joihin tekoälyratkaisu vaikuttaa.
- Yhteiskehittäminen on käytäntö, jossa sidosryhmät, mukaan lukien loppukäyttäjät tai asiakkaat, tekevät yhteistyötä suunnitteluprosessin aikana jakamaan näkemyksiä eri rooleissa ja eri osaamisalueilla toimivien sidosryhmien välillä. Näin voidaan paremmin ymmärtää loppukäyttäjien tarpeita ja suunnitella ratkaisuja, joissa otetaan huomioon eri käyttäjät.

(J. Morley et al., 2019) / (UNESCO, 2023) / (Interaction Design Foundation, Co-creation) / (Anika Sanin, 2020)

Sidosryhmien osallistuminen tekoälyratkaisujen kehittämiseen

Kehityksen sidosryhmät



Kehittävä organisaatio

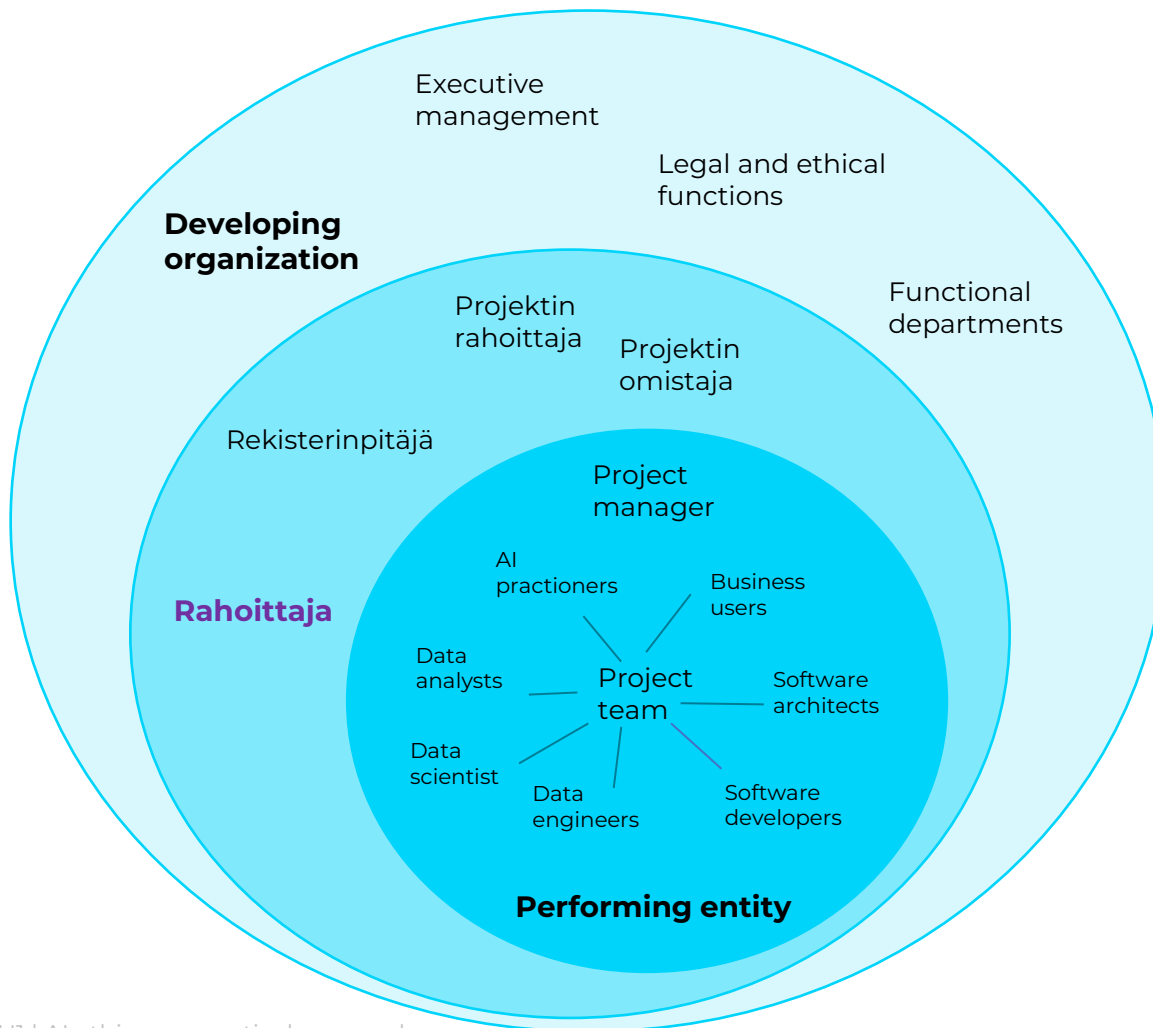
Sponsoroi tekoälyprojektia ja määrittelee projektin laajuuden.

Projektin roolit ovat johtoryhmä, toiminnalliset osastot sekä laki- ja eettiset toiminnot:

- Johtoryhmä vastaa yrityksen strategisista, taloudellisista ja kehityspäätöksistä.
- Toiminnallinen osasto on yrityksen sisäinen erikoistunut organisaatiosyksikkö, joka keskittyy tiettyihin tehtäviin tai toimintoihin, jotka on usein ryhmitelty yhteisten toimintojen, kuten talouden, markkinoinnin tai henkilöstöhallinnon, perusteella.
- Organisaation laki- ja eettiset toiminnot sisältävät erilaisia rooleja ja vastuita, joilla varmistetaan lakien, määräysten ja eettisten normien noudattaminen. (G.J. Miller, 2022)

Sidosryhmien osallistuminen tekoälyratkaisujen kehittämiseen

Kehityksen sidosryhmät



Rahoittaja

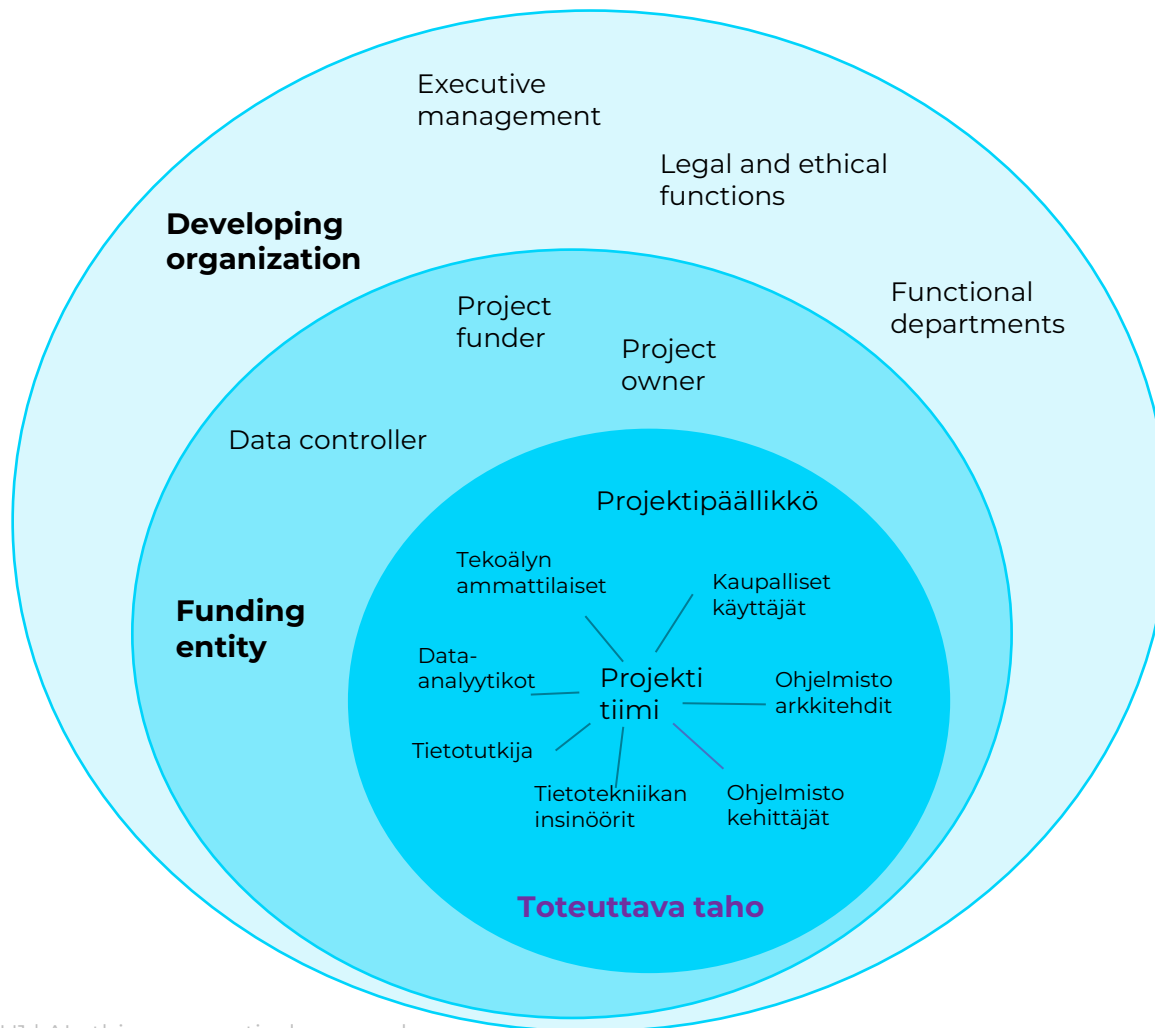
Projektin roolit ovat projektin rahoittaja, projektin omistaja ja tietojen valvoja.

- Projektin rahoittaja rahoittaa projektin ja osoittaa sille resurssit.
- Projektin omistaja tarjoaa strategisen ohjauksen.
- Tietojen valvoja vastaa tietojen hallinnasta ja niiden käytöstä järjestelmän kehittämisessä.

(G.J. Miller, 2022)

Sidosryhmien osallistuminen tekoälyratkaisujen kehittämiseen

Kehityksen sidosryhmät



Toteuttava tah

Projektin roolit ovat projektipäällikkö ja projektitiimi.

- Projektipäällikkö vastaa projektin tuloksista ja varmistaa, että eettisiä, tietosuoja- ja turvallisuusstandardeja sekä odotuksia noudatetaan ja että asiaankuuluvat sidosryhmät ovat mukana.
- Tekoälyprojektin tiimi koostuu ohjelmistoarkkitehdeistä, ohjelmistokehittäjistä, data-insinööreistä, data-analyytikoista, tekoälyn ammattilaisista ja liiketoiminnan käyttäjistä.

(G.J. Miller, 2022)

Sidosryhmien osallistuminen tekoälyratkaisujen kehittämiseen

Käyttäjät



Toimintayksikkö

Operatiivinen organisaatio

- Operatiivinen organisaatio on asiakkaan sidosryhmä.
- Kehitysyhteisön tavoin operatiivisen organisaation projektitehtäviin kuuluvat johtotehtävät sekä oikeudelliset ja eettiset tehtävät.
- Operatiivisella tasolla projektitehtäviä hoitavat loppukäyttäjät, mallien ylläpitäjät, rekisterinpitäjät ja alustan omistajat.
- Loppukäyttäjät ovat henkilöitä tai ryhmiä, jotka ovat tekemisissä tekoälyjärjestelmän kanssa työ- tai palvelusopimuksen perusteella operatiivisen organisaation kanssa tai kuluttajina.
- Mallien ylläpitäjät vastaavat tekoälyjärjestelmässä käytettävien koneoppimismallien jatkuvasta hallinnasta, päivityksistä ja optimoinnista.
- Rekisterinpitäjät määrittelevät henkilötietojen käsittelyn tarkoitukset ja keinot tekoälyjärjestelmässä.
- Alustan omistajat vastaavat tekoälyjärjestelmän kokonaisinfrastruktuurista ja käyttöönotosta.

Sidosryhmien osallistuminen tekoälyratkaisujen kehittämiseen

Ulkoiset sidosryhmät



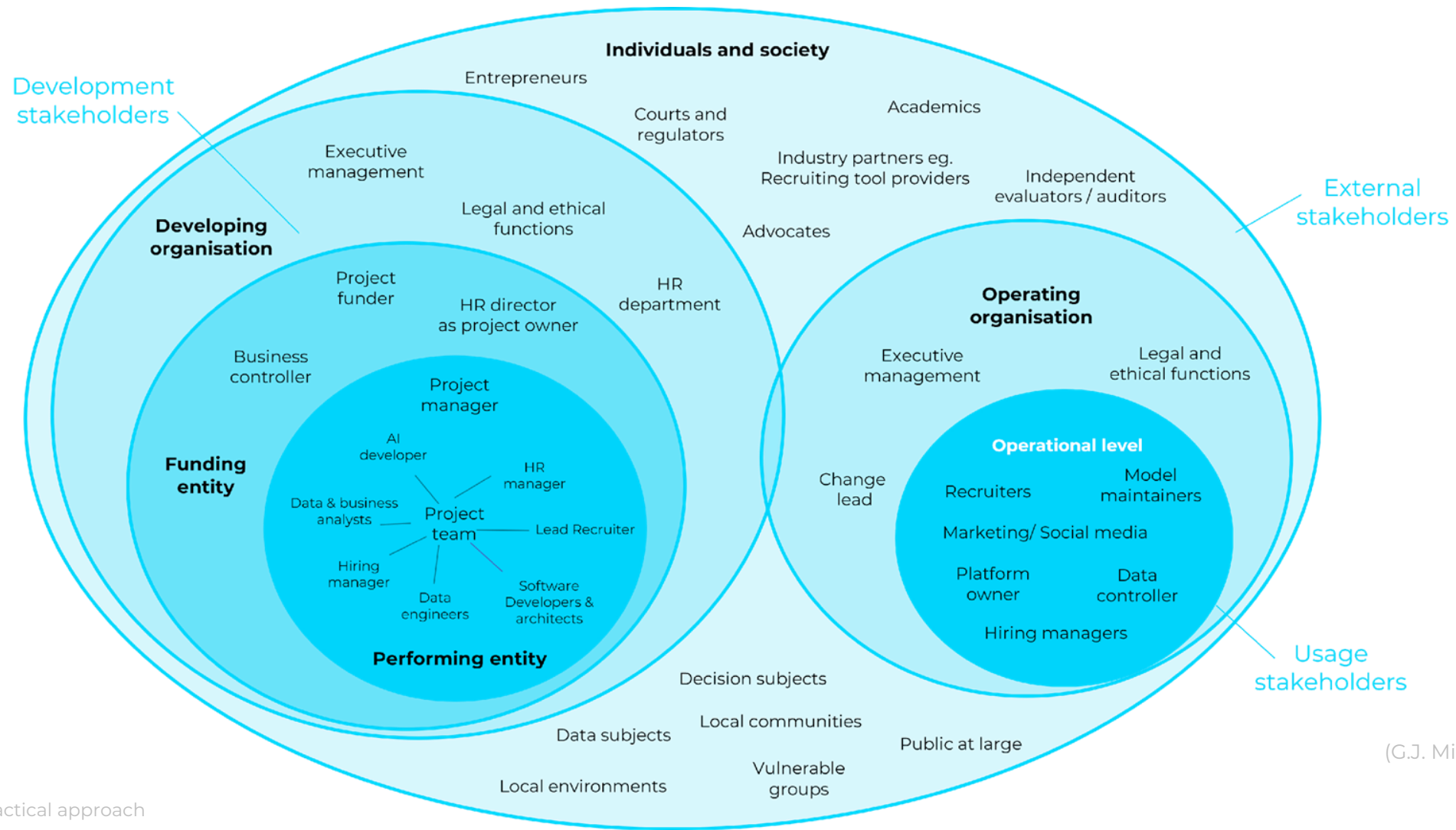
Yksilöt ja yhteiskunta

Yksilöt ovat henkilöitä, joilla on sopimuksia kehittäjän tai käyttäjän kanssa, kuten palvelu- tai työsopimus. Tähän ryhmään kuuluvat rekisteröidyt, päätöksenteon kohteet ja työntekijät, ja heitä pidetään passiivisina sidosryhminä. Yhteiskunta koostuu yksilöistä ja yleisöstä. Tekoälyjärjestelmän kehittäminen tai käyttö voi vaikuttaa heihin kielteisesti.

Tähän ryhmään kuuluvat paikallisyhteisöt, ympäristöt ja haavoittuvat ryhmät, kuten vammaiset, vähemmistöt, alaikäiset ja suuri yleisö.

- Tekoälyinsinöörit ovat osa laajempaa yhteiskunnallista keskustelua, jossa heidän on aktiivisesti integroitava muita näkökulmia.
- Edustajat ovat ryhmiä ja organisaatioita, joihin tekoälyjärjestelmä ei vaikuta, mutta jotka voivat edustaa muita projektissa.
- Virallisia edustajia ovat tuomioistuimet, lait, sääntelyviranomaiset ja ammattiliitot.
- Yrittäjät voivat olla mukana tekoälyjärjestelmän kehittämisessä.
- Riippumattomat arvioijat ovat ulkopuolisia henkilöitä tai ryhmiä, kuten toimittajia, tilintarkastajia tai arvioijia.

Sidosryhmien osallistuminen tekoälyratkaisujen kehittämiseen Stakeholders in AI development



EETTISET PERIAATTEET JA SIDOSRYHMIEN OSALLISTUMINEN

tekoälyn kehityksen
elinkaareen



Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen



Tekoälyn elinkaaren aikana on tärkeää keskittyä eettisiin näkökohtiin ja sidosryhmien osallistamiseen.

Seuraavilla dioilla käydään läpi elinkaari kolmesta näkökulmasta:

- 1. Vaihekuvaus**
- 2. Tärkeimmät eettiset periaatteet**
- 3. Osallistuvat sidosryhmät**

Kaikki eettiset periaatteet eivät ole yhtä tärkeitä kaikissa tilanteissa tai kehityksen kaikissa vaiheissa, joten seuraavassa esitetään vain tärkeimmät näkökohdat ja näkökulmat.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Ongelman määrittely ja liiketoiminnan ymmärtäminen



Ratkaiseva ensimmäinen vaihe onnistuneen projektin toteuttamiseksi

- Ymmärrä ratkaistava ongelma tai tutkittava mahdollisuus.
- Suorita tutkimusta ja sidosryhmähaastatteluja selkeän ongelmanmäärittelyn laatimiseksi.
- Ymmärrä asiakkaiden tarpeet ja miten ne voidaan muuttaa toteuttamiskelpoiseksi liiketoimintaratkaisuksi.
- Määritä ratkaisun tavoitteet ja menestyskriteerit.
- Suorita toteutettavuustutkimus sen selvittämiseksi, onko tekoäly sopivin ratkaisu.
- Arvioi projektin tekninen ja eettinen toteutettavuus.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Ongelman määrittely ja liiketoiminnan ymmärtäminen

Huomioon otettavat eettiset periaatteet



Perustan luominen – käsittele eettisiä perusasioita, kuten ihmisoikeudet, ympäristövaikutukset ja yhteiskunnalliset huolenaiheet.

Hyvän tekeminen – varmista, että tekoälyratkaisu parantaa yksilöiden hyvinvointia ja yhteiskunnan hyvinvointia.

Tärkeimmät huomioon otettavat alueet:

Sidosryhmien osallistuminen: Luo luotettavia järjestelmiä, jotka tukevat ihmisten hyvinvointia. Ota vaikutuksen kohteena olevat henkilöt mukaan kehitysprosessiin.

Perusoikeuksien suojele: Turvaa yksilön oikeudet ja vahvista eettistä integriteettiä.

Ympäristön kestävyys: Minimoi ympäristövaikutukset kestävien käytäntöjen avulla.

Perustelut: Määritä järjestelmän tarkoitus selkeästi ja yhdistä se konkreettisiin yhteiskunnallisiin hyötyihin.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Ongelman määrittely ja liiketoiminnan ymmärtäminen

Huomioon otettavat eettiset periaatteet



Haitan välttäminen

- Varmista, että ongelman tunnistaminen ja ehdotettu ratkaisu eivät vahingoita yksilöitä tai yhteisöjä.
- Harkitse huolellisesti ja seuraa järjestelmän vaikutuksia fyysiseen ja henkiseen hyvinvointiin.
- Arvioi mahdolliset turvallisuusriskit ja niiden vähentämisstrategiat.

Oikeudenmukaisuus

- Arvioi järjestelmän vaikutuksia instituutioihin, demokratiaan ja yhteiskuntaan.
- Harkitse, suositaanko tai syrjitäänkö tiettyjä ryhmiä.
- Tutki, voiko projekti loukata yksilön autonomiaa.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Ongelman määrittely ja liiketoiminnan ymmärtäminen Sidosryhmät



- **Loppukäyttäjät:** tekoälyratkaisun ensisijaiset käyttäjät
- **Tekoäly-/datatieteilijät:** muotoilevat ongelman ja konseptoivat ratkaisun
- **Vaikutusalueet**
- **Alan asiantuntijat:** syvälinen tuntemus alasta, jolla tekoälyä sovelletaan
- **Tekoälyn kehittäjät:** varhainen osallistuminen antaa heille syvällisen ymmärryksen ongelmasta
- **Projektipäälliköt:** valvovat koko projektin etenemistä
- **Kaupalliset sidosryhmät:** varmistavat yhdenmukaisuuden liiketoimintatavoitteiden ja resurssien kanssa
- **Eettiset ja yhteiskuntatieteilijät:** tuovat esiin mahdolliset vaikutukset

(D. De Silva & D. Alahakoon, 2022)

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Suunnitteluvaihe



Tässä vaiheessa ongelmanmäärittelyvaiheessa tehdyt pohjatyt muutetaan yksityiskohtaisiksi suunnitelmiksi tekoälyjärjestelmän rakentamiseksi ja käyttöönottamiseksi

- Määritä projektin tavoitteet (tavoitteet ja tulokset).
- Arvioi taloudelliset kustannukset suhteessa odotettuihin hyötyihin.
- Luo luonnos tekoälyjärjestelmän välttämättömistä toiminnoista.
- Päätä lähestymistapa tietojen hyödyntämiseen, kuten valmiiden mallien käyttöön.
- Määritä mittarit projektin onnistumisen arvioimiseksi.
- Sisällytä eettiset ohjeet kehitystyön ohjaamiseksi.
- Laadi aikataulu ja hallinnoi sidosryhmien osallistumista.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Suunnitteluvaihe

Huomioon otettavat eettiset periaatteet



Oikeudenmukaisuus | Epäoikeudenmukaisen vinouman välttäminen



Suunnitteluvaatimukset

- Ota käyttöön mekanismit, joilla tunnistetaan ja lievennetään dataan ja algoritmeihin liittyviä vinoumia.
- Pyrkikää eri ryhmien välisiin tasapuolisiin tuloksiin.

Kehittäjien tietoisuus

- Varmistakaa, että kehittäjät ovat tietoisia mahdollisista vinoumista, jotka liittyvät erityisesti esimerkiksi rotuun, ihonväriin, syntyperään, sukupuoleen, ikään, kieleen, uskontoon, poliittiseen mielipiteeseen, kansalliseen, etniseen tai sosiaaliseen alkuperään, taloudelliseen asemaan, syntymäolosuhteisiin tai vammaisuuteen.

Viestintä ja reagointi

- Laadi selkeät menettelytavat vinoumien, kielteisten vaikutusten tai kompromissien viestittämiseksi, raportoimiseksi ja niihin reagoimiseksi.

Tietojen hallinta

- Harkitse tietojen tallennusjärjestelmiä, joilla estetään tietojen eristäminen ja siitä johtuvat vinoumat.

Riskienhallinta

- Laadi "mitä tapahtuu, jos" -toimenpiteitä odottamattomien tulosten varalta.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Suunnitteluvaihe

Huomioon otettavat eettiset periaatteet



Selitetävyys / läpinäkyvyys

- Varmistetaan tekoälyn päätösten selkeä dokumentointi ja käyttäjäystävälliset selitykset.
- Varmistetaan tietolähteiden ja algoritmisten prosessien läpinäkyvyys.

Hyvän tekeminen

- Sidosryhmät otetaan mukaan kehitystyöhön, jotta voidaan ottaa huomioon erilaiset eettiset näkökohdat.

Autonomia

- Tekoälyratkaisuihin otetaan käyttöön ihmisen suorittamia valvontamekanismeja, joilla varmistetaan inhimillisten arvojen säilyminen.

HLEG/ Unesco / Morley / Prem

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Suunnitteluvaihe Sidosryhmät



- Ohjelmistoarkkitehdit: luovat tekoälyratkaisun teknisen perustan ottaen huomioon sekä nykyiset että tulevat tarpeet.
- Käyttäjäkokemuksen (UX) suunnittelijat: ratkaisevassa asemassa luotaessa käyttöliittymiä, jotka parantavat käyttäjien hyväksyntää ja tyytyväisyyttä ja varmistavat, että käyttäjät voivat olla vuorovaikutuksessa tekoälyjärjestelmän kanssa tehokkaasti.
- Eettiset asiantuntijat: käsittelevät eettisiä kysymyksiä.
- Tekoälyn ammattilaiset: tekevät yhteistyötä UX-suunnittelijoiden kanssa tekoälykomponenttien saumattoman integroimiseksi.
- Oikeudelliset ja eettiset tehtävät: lieventävät oikeudellisia riskejä ottamalla huomioon vaatimustenmukaisuusvaatimukset jo suunnitteluvaiheessa ja auttavat suunnittelemaan ratkaisuja, jotka noudattavat tietosuoja- ja yksityisyydensuojamääräyksiä.
- Tekoälymallin vaikutukset ihmisiin ja yhteisöihin: osallistava suunnittelu.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Tietojen kerääminen, ymmärtäminen ja valmistelu



Kun tavoite ja sen saavuttamissuunnitelma on vahvistettu, seuraava vaihe on työskennellä asiaankuuluvien tietojen parissa.

- Kerää kaikki projektiin tarvittavat tiedot.
- Kuvaile, tutki, tarkista, valitse, puhdista, rakenna, integroi ja muotoile tiedot sopiviin luokkiin ja käsitteisiin käyttötarkoitusta varten.
- Käsittele ja ratkaise puuttuviin tietoihin liittyvät ongelmat, jotta tiedot ovat täydelliset.
- Tunnista ja ymmärrä olemassa olevien tietojen taustalla olevat oletukset.
- Selvitä, miten tietokokonaisuuksia käytetään päätöksenteossa.
- Pidä tietojen laatu korkeana, erityisesti jos tiedot ovat epäselviä tai puutteellisia.

Tämä on aikaa vievä vaihe, jossa rakennetaan toimivan ratkaisun perusta. Ratkaisun luotettavuus ja suorituskyky ovat suoraan verrannollisia sen perustana olevien tietojen laatuun.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Tietojen kerääminen, ymmärtäminen ja valmistelu

Huomioon otettavat eettiset periaatteet



Haitan välttäminen 1/2

- Tietojen eheys ja laatu, jotka aiheuttavat vinoumaa
- Huomioi mahdolliset vinoumat, jotka suosivat tiettyjä väestöryhmiä. Varmista tarkkuus yhdistettäessä tietokantoja (esim. samannimisiä henkilöitä ei saa sekoittaa).
- Suorita perusteellinen due diligence -tarkastus tietojen tarkkuuden arvioimiseksi ja varmistamiseksi, jotta ne eivät vahvista epäoikeudenmukaisia vinoumia.
- Vältä käyttämästä tiettyihin väestöryhmiin perustuvia erillisiä tietokantoja, jotka voivat vahvistaa vinoumia.
- Ota käyttöön yhtenäinen tietovarasto, kuten tietovarastokeskus, vinouman minimoimiseksi (huomautus: kallis).



Voiko vinouma olla merkityksellinen tekijä tekoälyratkaisussa? Esimerkiksi autovakuutuksiin vaikuttavat onnettomuustyyppit ja onnettomuuksiin osalliset demografiset tekijät. Oikeudelliset rajoitukset: Sukupuoli ei voi määrittää autovakuutusten hintoja Euroopan unionin tuomioistuimen vuonna 2012 antaman tuomion mukaisesti.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Tietojen kerääminen, ymmärtäminen ja valmistelu

Periaatteet



Haitan välttäminen 2/2

Tietosuoja ja luottamuksellisuus:

- Huomioi tietosuojaan ja luottamuksellisuuteen liittyvät tyypilliset riskit tietojen keräämisen aikana.
- Varmista, että tekoälyjärjestelmät noudattavat tietosuojaa koko elinkaaren ajan ja noudattavat GDPR:n kaltaisia säännöksiä.

Hyökkäysten kestävyys ja turvallisuus:

- Suojaa tietojoukkoja hyökkäyksiltä ja puutu mahdollisiin turvallisuusriskeihin vahvan tietoturvan ylläpitämiseksi.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Tietojen kerääminen, ymmärtäminen ja valmistelu

Periaatteet



- Tietotieteilijät löytävät tiedoista malleja ja trendejä, jotka muuttuvat tiedoksi.
- Tekoälyinsinöörit: vastaavat tietojen arvioinnista ja tarkistamisesta.
- Tietojen analysoijat auttavat ymmärtämään tietojen ominaisuuksia ja tukevat tietojen valmistelustrategioita.
- Tietotekniikka-insinöörit: vastaavat tietojen valmistelusta ja laadunvarmistuksesta.
- Tietosuojavastaava (DPO): vastaa siitä, että organisaatio noudattaa sääntöjä käsitellessään työntekijöiden, asiakkaiden, toimittajien tai muiden henkilöiden (rekisteröityjen) henkilötietoja.
- Eettiset asiantuntijat: käsittelevät eettisiä näkökohtia tietojen käsittelyssä ja valinnassa.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Mallien kehittäminen ja koulutus



Tässä vaiheessa kehitetään tekoälymalli määritellyn ongelman ratkaisemiseksi. Tähän sisältyy useita kriittisiä vaiheita:

- Valitse ja määritä algoritmimalli ja työkalut sekä määritä niiden käyttö etukäteen.
- Luettele ja arvioi valittujen työkalujen vaikutukset, ottaen erityisesti huomioon mahdolliset haitat ja kompromissit.
- Päätä, mitkä mallin elementit tulisi korostaa ja mitkä voivat jäädä toissijaisiksi tai jopa piiloon.
- Suorita useita tarkennuskierroksia koulutuksen aikana ja paranna mallia jatkuvasti.
- Työkalujen valinta edellyttää usein ristiriitaisten etujen tasapainottamista. Esimerkiksi on päätettävä, onko parempi ottaa riski, että roskapostia päätyy oikeaan postilaatikkoon, vai että oikeita sähköposteja päätyy roskapostikansioon.

Tämä vaihe on ratkaiseva, koska se vaikuttaa suoraan tekoälyratkaisun tehokkuuteen ja vaikuttavuuteen.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Mallien kehittäminen ja koulutus

Huomioon otettavat eettiset periaatteet



Selitettävyyys

- Teknisiin näkökohtiin keskittymisen lisäksi on tärkeää tarkistaa tekoälyjärjestelmän selitettävyyys ja mahdolliset vinoumat, jotta niihin voidaan puuttua ajoissa.
- Kehitysprosessin läpinäkyvyys on varmistettava, jotta sidosryhmät voivat ymmärtää sen. Tähän kuuluu tietolähteiden, mallivalintojen ja päätösten perustelujen selkeä dokumentointi.

Oikeudenmukaisuus

- Tehokkuuden ja muiden vaatimusten väliset kompromissit on tunnistettava ja raportoitava avoimesti.
- Luo prosessi, jolla tunnistetaan ja arvioidaan näiden kompromissien vaikutukset.
- Tekoälyinsinöörien on perusteltava päätöksensä ottaen huomioon mahdolliset kielteiset vaikutukset niihin henkilöihin, joita päätökset koskevat.

Morley (2019) Prem (2023)

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Mallien kehittäminen ja koulutus Sidosryhmät



- Tekoälyn/koneoppimisen tutkijat muuntavat ongelman määritelmän prototyyppiseksi tekoälymalliksi.
- Tekoälyinsinööreillä on asiantuntemusta valita tiettyihin tarkoituksiin sopivimmat algoritmiset työkalut.
- Data-analyytikot ovat keskeisessä asemassa koneoppimismallien kehittämisessä ja optimoinnissa.
- Eettiset asiantuntijat ovat välttämättömiä eettisten periaatteiden integroimiseksi kehitysprosessiin, jotta vältetään tahattomat seuraukset.
- Kaupalliset käyttäjät varmistavat, että mallit täyttävät liiketoiminnan vaatimukset ja odotukset.

(D. De Silva & D. Alahakoon, 2022), (J. Rochel & F. Évequoz, 2020)

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Mallien kehittäminen ja koulutus



Kun malli on kehitetty ja koulutettu, sen suorituskyky on testattava ja arvioitava:

- Arvioi malli ennalta määriteltyjen tavoitteiden ja menestyskriteerien perusteella.
- Ilmoita selkeästi taustalla olevat oletukset ja standardit, joihin tietojoukot perustuvat.
- Suorita testejä käyttämällä näkemättömiä tietoja tai prosessin vaiheita varmistaaksesi mallin kestävyys.
- Jos suorituskyky ei ole tyydyttävä, harkitse mallin parametrien, arkkitehtuurin tai jopa tietomallin taustalla olevien tietojoukkojen muuttamista.
- Kaikki puutteet on keskusteltava avoimesti ja rehellisesti projektin johtajan ja muiden sidosryhmien kanssa, koska ne voivat vaikuttaa merkittävästi yksilöihin.
- Sovitaan, jatketaanko toistojen tekemistä vai siirrytäänkö käyttöönottovaiheeseen.

Tämä vaihe on ratkaisevan tärkeä mallin luotettavuuden ja tehokkuuden varmistamiseksi ennen sen tuotantoon siirtämistä.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Mallien kehittäminen ja koulutus

Huomioon otettavat eettiset periaatteet



Oikeudenmukaisuus / Hyvän tekeminen / Haitan välttäminen

Eettiset auditoinnit ja mittarit:

- Varmista eettisten periaatteiden, kuten oikeudenmukaisuuden, hyvän tekemisen ja haitan välttämisen, noudattaminen auditoinneilla ja auditoitavuutta varten suunnitelluilla mittareilla.
- Turvallisuus ja sietokyky:
- Testaa sietokyky hyökkäyksiä vastaan ja laadi varasuunnitelma järjestelmän turvallisuuden varmistamiseksi.
- Negatiivisten vaikutusten ja mahdollisten riskien käsittely:
- Toteuta strategioita negatiivisten ongelmien minimoimiseksi ja raportoimiseksi.
- Laadi korjausmekanismi mahdollisten haittojen varalta, jos jokin menee pieleen.
- Laadi riskinarviointitekijät, joissa keskitytään yksityisyyteen, sosiaalisiin vaikutuksiin ja vinoumiin.
- Selitettävyys:
- Selitä selkeästi sekä tekniset prosessit että mallin kehittämiseen liittyvät inhimilliset päätökset.

Huomaa, että ratkaisujen käyttöönottoon kiireessä olevat insinöörit voivat olla alttiimpia jättämään huomiotta kriittisiä asioita.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Mallien kehittäminen ja koulutus

Sidosryhmät



- Tekoälyn insinöörit tulkitsevat mallinnuksen tulokset.
- Laadunvarmistustiimit varmistavat, että mallit täyttävät laatustandardit ja ovat virheettömiä eivätkä aiheuta tahattomia seurauksia.
- Eettiset asiantuntijat varmistavat tekoälymallien eettisen käytön.
- Käyttäjien edustajat varmistavat, että tekoälymallit vastaavat käyttäjien odotuksia ja tarpeita.
- Lakiasiain- ja compliance-tiimit vähentävät oikeudellisia riskejä varmistamalla, että asiaankuuluvia säännöksiä noudatetaan.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Käyttöönotto



Kun tekoälyratkaisu on viimeistelty tyydyttäväksi, se otetaan tuotantokäyttöön ratkaisemaan todellisia ongelmia. Tämän vaiheen monimutkaisuus riippuu ratkaisusta itsestään.

- Alun perin ratkaisu otetaan yleensä käyttöön pienemmälle asiantuntija- ja käyttäjäryhmälle.
- Integroi malli olemassa oleviin järjestelmiin, luo malliin perustuva sovellus tai palvelu tai hyödynnä saatuja tietoja offline-kontekstissa, esimerkiksi johdolle laadittavassa raportissa.
- Luo palautesilmukka suorituskyvyn ja tulosten seurantaan varten ja toista tarvittaessa.
- Ota käyttöön riskienhallintastrategiat mahdollisten ongelmien ratkaisemiseksi käyttöönoton aikana.
- Suunnittele jatkuva seuranta ja ylläpito operatiivisessa vaiheessa.
- Tarkista koko projekti ja laadi lopullinen raportti, jossa dokumentoidaan prosessi ja tulokset.

Tämä jäsennelty lähestymistapa varmistaa, että tekoälyratkaisu integroidaan tehokkaasti ja että se toimii tarkoitetulla tavalla todellisissa tilanteissa.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Käyttöönotto

Huomioon otettavat eettiset periaatteet



Autonomia

Ihmisen toiminta

- Varmistetaan, että käyttäjät ovat tietoisia siitä, kun päätökset, sisältö, neuvot tai tulokset ovat tekoälyn tuottamia. Korostetaan vuorovaikutusta tekoälyratkaisun kanssa, jotta vältetään liiallinen riippuvuus siitä.

Ihmisen valvonta

- Varmistetaan, että ihmisen valvonta on käytössä, jotta voidaan puuttua tarvittaessa ja varmistaa, että päätökset voidaan tarkistaa tai muuttaa ihmisten toimesta.

Oikeudenmukaisuus

- Otetaan käyttöön tehokkaat tarkastustoimenpiteet järjestelmän suorituskyvyn ja eettisten periaatteiden noudattamisen säännöllistä arviointia varten.
- Luo mekanismit, joiden avulla käyttäjät voivat valittaa päätöksistä tai tehdä valituksia, ja varmistaa, että ilmiantajille on tarjolla suojelua.

Unesco (2023) / Morley (2019)

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Käyttöönotto

Huomioon otettavat eettiset periaatteet



Selitettävyys

- Korosta, että on tärkeää selittää käyttäjille sekä tekniset prosessit että niihin liittyvät ihmisten päätökset.
- Yritä yksinkertaistaa monimutkaista ihmisten käyttäytymistä helposti ymmärrettäviksi työkaluiksi välttämällä teknistä jargonia → tämä lisää luottamusta ja järjestelmän hyväksyntää.
- Pidä tekoälyn toiminnot ja rajoitukset avoimina.

Pitkän aikavälin vaikutusten arviointi

- Arvioi tässä vaiheessa pitkän aikavälin vaikutukset yhteiskuntaan ja ympäristöön.
- Harkitse tarvittavia parannuksia, joilla varmistetaan, että tekoälyjärjestelmä vaikuttaa positiivisesti ajan mittaan.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Käyttöönotto Sidosryhmät



- Tekoälyn/koneoppimisen insinöörit muuntavat prototyyppisen tekoälymallin käyttöönotettavaksi palveluksi tai ratkaisuksi, joka on kaikkien sidosryhmien ja loppukäyttäjien saatavilla.
- Turvallisuusasiantuntijat suojaavat käyttöönotetun järjestelmän kyberturvallisuusriskeiltä.
- Loppukäyttäjät tunnistavat ongelmat tai rajoitukset, jotka eivät ole tulleet esiin kontrolloiduissa ympäristöissä suoritetuissa testeissä.
- Käyttäjien edustajat keräävät käyttäjiltä palautetta käyttöönotetun tekoälyratkaisun käytettävyydestä ja suorituskyvystä ja varmistavat, että käyttöönotettu ratkaisu vastaa käyttäjien odotuksia ja tarpeita.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Valvonta ja käyttö



Käyttöönoton jälkeen tekoälyratkaisu vaatii jatkuvaa ylläpitoa, päivityksiä ja jatkuvaa arviointia.

- Varmista, että malli toimii odotetusti seuraamalla sen tehokkuutta säännöllisesti.
- Seuraa jatkuvasti loppukäyttäjien toimintaa kerätäksesi toiminnallisia näkemyksiä ja havaitaaksesi mahdolliset ongelmat.
- Päivitä malli säännöllisesti uusimmilla tiedoilla sen tarkkuuden ja relevanssin ylläpitämiseksi.
- Tarkenna mallia käyttäjien palautteen perusteella toiminnallisuuden ja suorituskyvyn parantamiseksi.
- Toistuvat päivitykset ovat usein tarpeen, ja niitä tulisi pitää tekoälyn elinkaaren normaalina osana, ei epäonnistumisena.
- Seuraa sijoitetun pääoman tuottoa ja muita relevantteja mittareita alun perin asetettujen tavoitteiden perusteella.

Tämä vaihe on kriittinen sen varmistamiseksi, että tekoälyratkaisu pysyy tehokkaana ja vastaa edelleen käyttäjien tarpeita.

De Silva & Alahakoon (2022) / Data Science Process Alliance

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Valvonta ja käyttö

Huomioon otettavat eettiset periaatteet



Eettisten periaatteiden seuranta

- Tarkista ja arvioi säännöllisesti tekoälyjärjestelmän eettisten periaatteiden noudattamista ottaen huomioon mahdolliset muutokset tietokannoissa ja mallien rakenteissa.

Sidosryhmien jatkuva osallistaminen

- Pidä yllä aktiivista vuorovaikutusta käyttäjien, organisaatioiden ja laajemman yhteiskunnan kanssa jatkuvan palautteen saamiseksi.
- Luo jäsennellyt mekanismit, joilla varmistetaan säännöllinen ja järjestelmällinen palautteen ja vastausten kerääminen.

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Valvonta ja käyttö

Sidosryhmät



- DevOps- ja IT-operaatiotiimit valvovat järjestelmän suorituskykyä ja varmistavat jatkuvan käytettävyyden.
- Turvallisuussertifioijat arvioivat kehitetyn tekoälyratkaisun turvallisuuden.
- Asiantuntijajaneelit, ohjauskomiteat tai sääntelyelimet tarkastavat projektin teknisiltä ja eettisiltä näkökohdilta.

(D. De Silva & D. Alahakoon, 2022) / (Immersant Data Solutions, 2023) / (Miller, 2022)

Eettiset periaatteet ja sidosryhmien osallistuminen tekoälyn kehityksen elinkaareen

Tekoälyn kehityksen elinkaari Työkalut

Eettisten näkökohtien huomioon ottamiseen käytettävissä olevat työkalut vaihtelevat paljon tekoälyn kehitysvaiheen mukaan. Monet lähestymistavat voivat olla päteviä kaikissa vaiheissa, mutta jotkut vain tietyissä kehitysvaiheissa. Alla oleva luokittelu on karkea kuva siitä, miten käytettävissä olevat työkalut voidaan jakaa suunnitteluvaiheisiin.



KIITOS



Universitat
de les Illes Balears

ISQe
ENGAGING PEOPLE



**Euroopan unionin
osarahoittama**

Euroopan unionin rahoittama. Esitetyt näkemykset ja mielipiteet ovat ainoastaan tämän tekstin laatijoiden näkemyksiä eivätkä välttämättä vastaa Euroopan unionin tai Euroopan koulutuksen ja kulttuurin toimeenpanovirasto (EACEA) kantaa. Euroopan unioni ja EACEA eivät ole vastuussa niistä.

2022-1-ES01-KA220-HED-000085257