



Eettisen tekoälyn mikrokurssi

OPAS

**Osaamiskokonaisuus 6 | Tekoälyn
etiikka käytännössä**

Hankkeen numero:
2022-1-ES01-KA220-HED-000085257

Miten tätä Opasta käytetään?

Tämä dokumentti on interaktiivinen.
Dokumentissa on linkkejä lisätietoihin.



Painike, joka vie sinut dokumentin alkuun. Tämä kuvake näkyy sivujen oikeassa yläkulmassa.



Aina kun näet tämän nuolen, se tarkoittaa, että kyseessä on **interaktiivinen väriteksti**, jota voit napsauttaa ja johon on liitetty ulkoinen linkki.

VASTUUVAPAUTUS: Huomaa, että emme voi taata ulkoisen sisällön, kuten videoiden, jatkuvaa saatavuutta, sillä niiden tekijät tai isäntäalustat voivat muuttaa tai poistaa niitä.

Sisältö

Klikkaa valikkoa

01. Johdanto

02. Tekoälyn eettisten vaikutusten ymmärtäminen

**03. Epäeettisten tekoälykäytäntöjen
tunnistaminen ja lieventäminen**

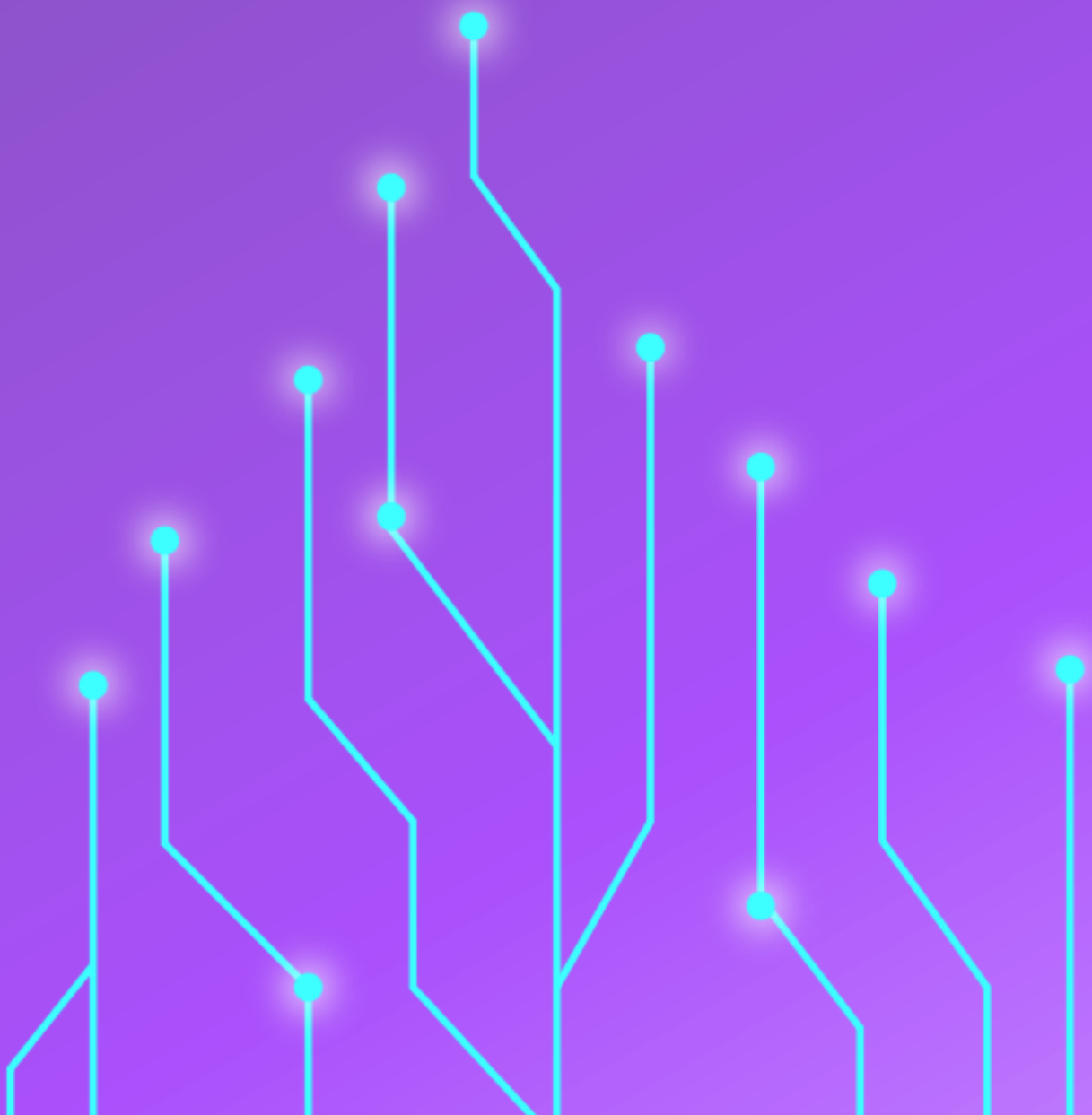
**04. Tekoälyn eettisten ohjeiden
käytännön soveltaminen**

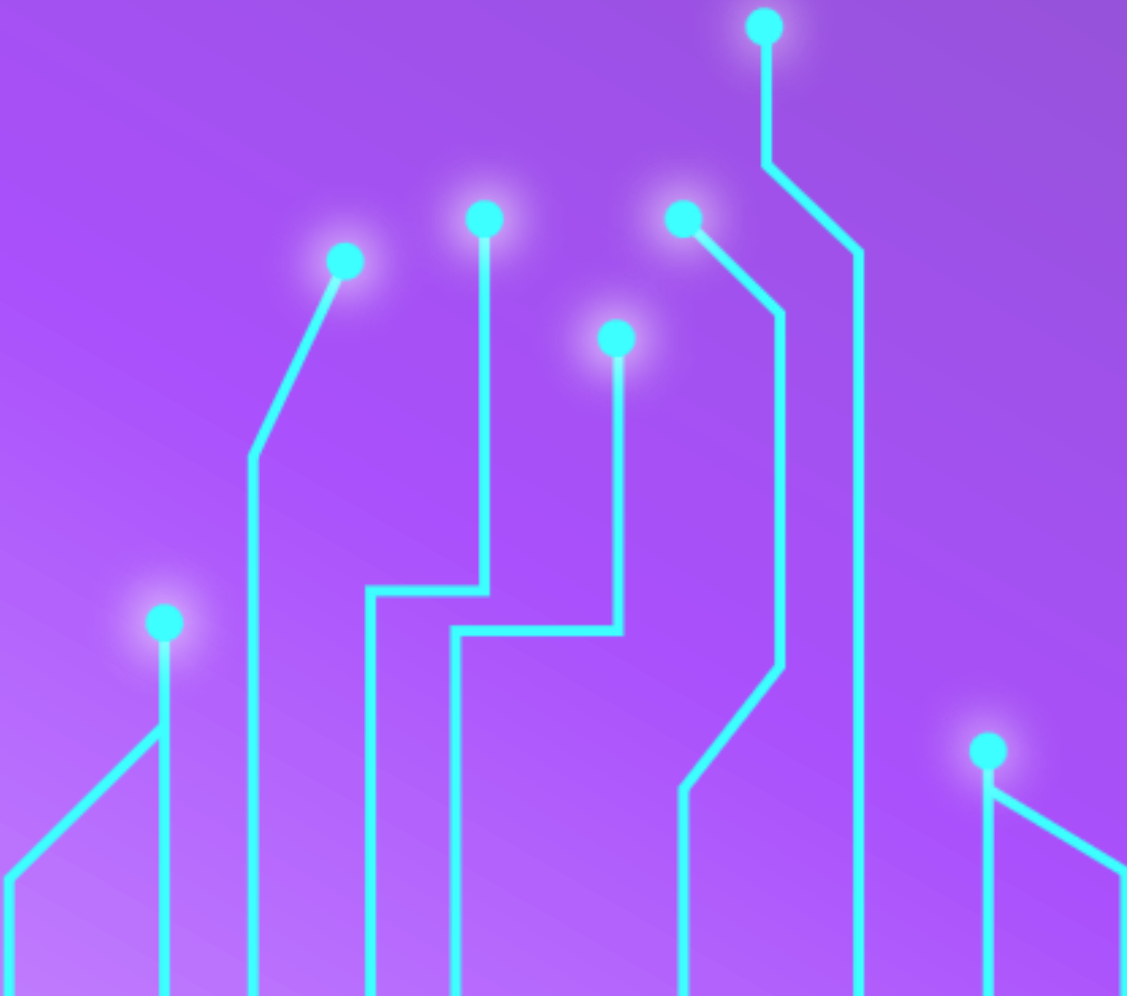
**05. Vastuullisen tekoälyn kehittämisen ja
käyttöönoton edistäminen**

06. Yhteenveto

01. Johdanto

Osaamiskokonaisuus 6 | Tekoälyn etiikka käytännössä





01. Johdanto

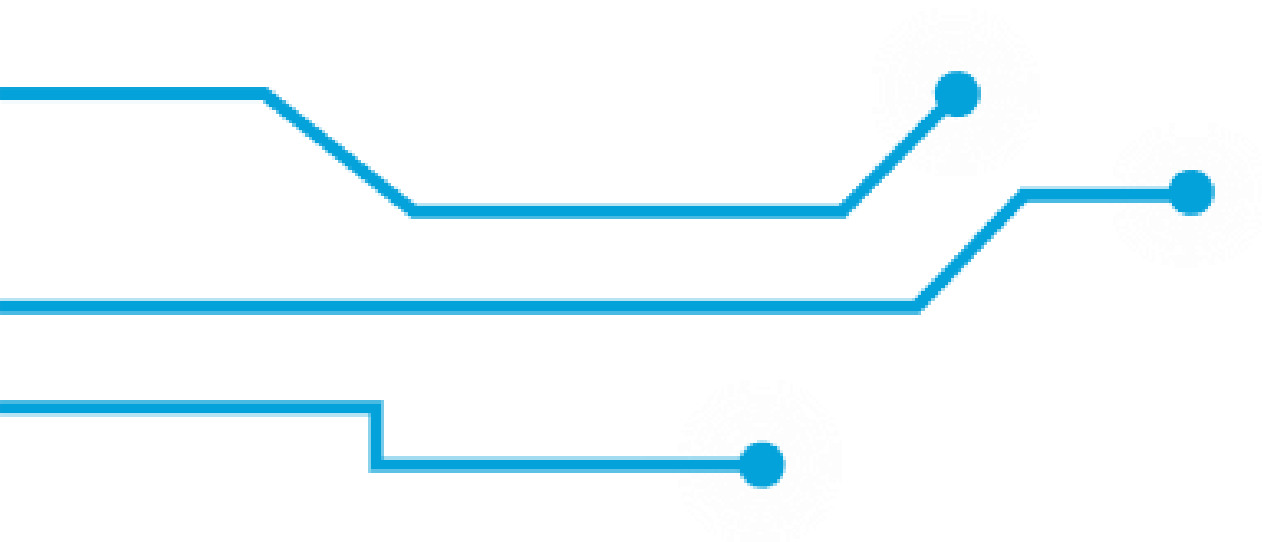
Tämä osaamiskokonaisuus on suunniteltu opiskelijoille, jotka haluavat syventyä tekoälyn etiikan käytännön näkökohtiin ja sisällyttää tekoälyn vastuullisen kehittämisen ja käytön periaatteet reaalimaailman skenaarioihin. Opiskelijoita kannustetaan toimimaan ennakoivasti eettisten ohjeiden täytäntöönpanossa tekoäly-ympäristöissä, hyödyntämään teoreettista tietoa ja muuntamaan se toimiviksi oivalluksiksi.

Tämän osaamiskokonaisuuden osaamistulokset:

- **Tekoälyn eettisten vaikutusten ymmärtäminen:** Opiskelijat saavat aluksi vankan perustan tekoälyn eri eettisistä ulottuvuuksista. Floridi et al. (2018) antamien, tekoälyn eettisiä huolenaiheita koskevien näkemysten mukaisesti opiskelijat tutustuvat tekoälyn mahdollisiin vaikutuksiin yhteiskuntaan, yksilöihin ja globaaleihin yhteisöihin.
- **Tekoälyn epäeettisten käytäntöjen tunnistaminen ja lieventäminen:** käsitellään tarvetta tunnistaa ja lieventää mahdollisia epäeettisiä käytäntöjä tekoälyn alalla. Opiskelijat oppivat, miten tunnistaa ja estää epäeettiset käytännöt tekoälyn kehityksessä ja käyttöönotossa, Bostromin (2014) pohdintojen pohjalta. Bostrom tutki tekoälyn tulevaisuutta ja sen mukauttamista inhimillisiin arvoihin.

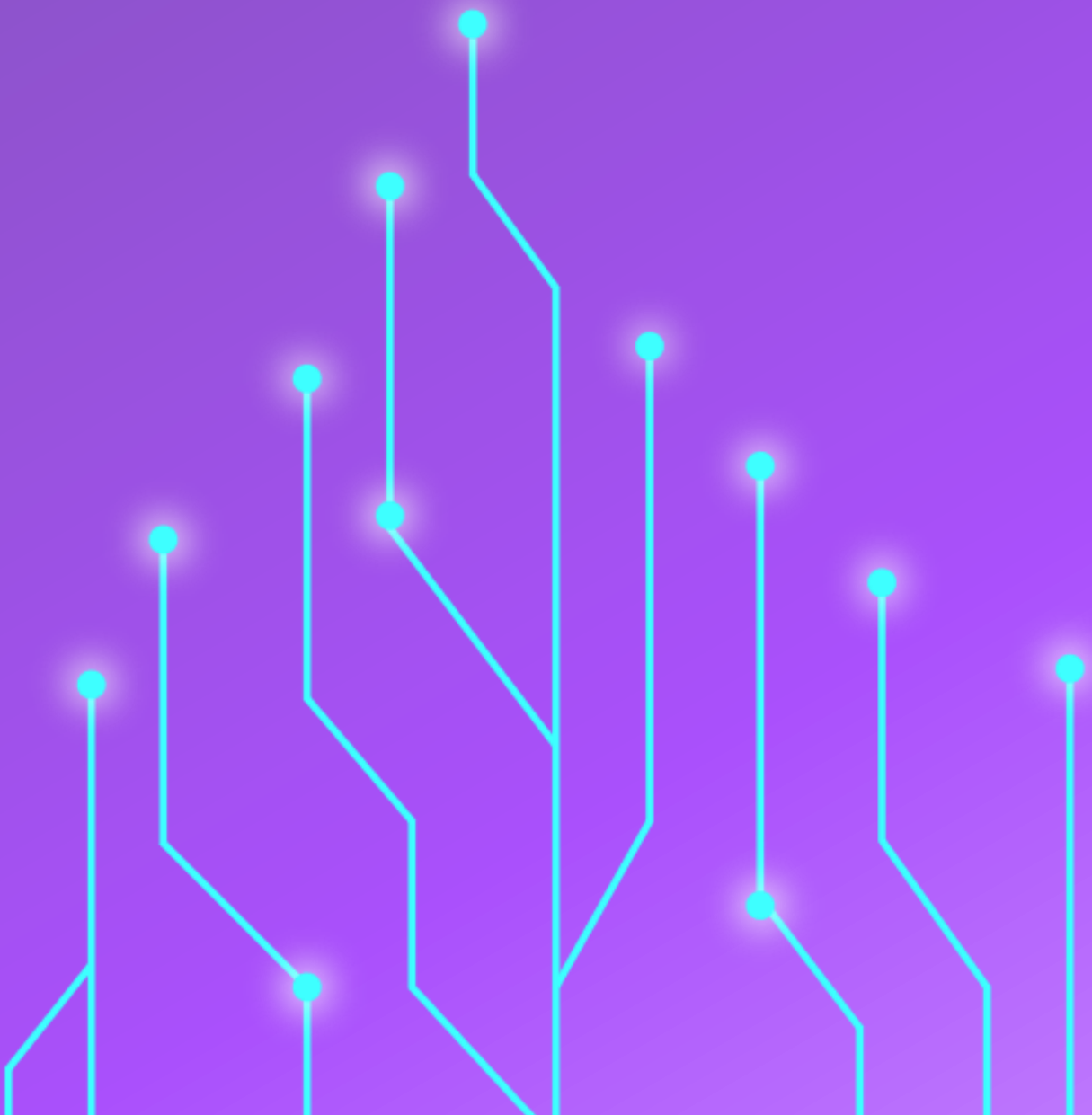


- **Tekoälyn eettisten ohjeiden käytännön soveltaminen:** tämän moduulin opiskelijoita kannustetaan siirtämään teoreettinen ymmärrys käytäntöön. Ryan & Stahlin (2020) teosten pohjalta, jotka keskittyvät eettisten ohjeiden laatimiseen, opiskelijat osallistuvat tehtäviin, jotka edistävät näiden ohjeiden käytännön soveltamista tekoälyn skenaarioissa.
- **Vastuullisen tekoälyn kehittämisen ja käyttöönoton edistäminen:** keskitytään sellaisen ympäristön edistämiseen, joka kannustaa vastuulliseen tekoälyn kehittämiseen ja käyttöönottoon. Jobin et al. (2019) kehyksen pohjalta, jossa käsitellään globaaleja näkökulmia tekoälyn etiikkaan, opiskelijat tutkivat menetelmiä, joilla edistetään globaalia yhteistyötä ja vastuullisen tekoälyn luomista.



02. Tekoälyn eettisten vaikutusten ymmärtäminen

Osaamiskokonaisuus 6 | Tekoälyn etiikka käytännössä





02. Tekoälyn eettisten vaikutusten ymmärtäminen

Kurssin aikana sukelsimme syvälle tekoälyn maailmaan ja tarkastelimme sen mahdollisuuksia ja siihen liittyviä riskejä. Nyt olet todennäköisesti jo hyvin tietoinen tekoälyn mullistavasta voimasta ja sen kasvavasta läsnäolosta eri aloilla. Tällaisen valtavan vallan myötä tulee kuitenkin myös syvälinen eettinen vastuu.

Eettisen tekoälyn moniulotteisessa maisemassa suunnistaessamme eri eettiset koulukunnat tarjoavat arvokasta opastusta. Esimerkiksi utilitarismi korostaa kokonaishyvinvoinnin maksimointia, mikä saa meidät pohtimaan, miten tekoäly voidaan suunnitella siten, että se hyödyttää mahdollisimman monia ihmisiä. Velvollisuusetiikassa taas keskitytään tekojen luontaiseen oikeellisuuteen tai vääryyteen ja kehoitetaan varmistamaan, että tekoälyjärjestelmät toimivat vakiintuneiden eettisten kehysten mukaisesti. Hyve-etiikka korostaa luonteen ja moraalisen kehityksen merkitystä ja kannustaa meitä edistämään vastuullisen tekoälyn kehittämisen kulttuuria, jossa ihmisarvot ovat etusijalla.

Tässä viimeisessä osaamiskokonaisuudessa tarkastelemme näitä keskeisiä ajattelutapoja ja siirrämme ne käytännön pohdinnoiksi eettisen tekoälyn rakentamiseksi. Tarkastelemme oikeudenmukaisuuteen, läpinäkyvyyteen ja vastuullisuuteen liittyviä kysymyksiä ja keskustelemme siitä, miten voidaan kehittää vankat puitteet, joilla varmistetaan, että tekoäly palvelee ihmiskuntaa vastuullisella ja hyödyllisellä tavalla.



Arvioimalla kriittisesti tekoälyn eettisiä vaikutuksia voimme tasoittaa tietä tulevaisuudelle, jossa nämä tehokkaat teknologiat edistävät oikeudenmukaisempaa ja tasapuolisempaa yhteiskuntaa.

> Yleiskatsaus eettisiin teorioihin ja niiden soveltamiseen tekoälyn alalla

Utilitarismi - merkittävä eettinen teoria, jossa korostetaan kokonaishyvinvoinnin tai "hyödyn" maksimoimista mahdollisimman monelle ihmiselle. Sovellettuna tekoälyyn tämä tarkoittaa, että etusijalle asetetaan koko yhteiskuntaa hyödyttävät tulokset ja samalla minimoidaan mahdolliset haitat.

- **Yksilöllinen vaikutus:** Kuvittele tekoälyllä toimiva turvakamerajärjestelmä naapurustossa. Vaikka se saattaisi parantaa asukkaiden turvallisuutta, se voisi myös lisätä valvontaa ja aiheuttaa tunteen siitä, että heitä valvotaan jatkuvasti. Utilitaristinen näkökulma punnitsisi yksilön oikeutta yksityisyyteen suhteessa koko yhteisön rikollisuuden mahdolliseen vähenemiseen.
- **Vaikutus yhteisöön:** Tekoälyllä toimivia algoritmeja käytetään yhä useammin sosiaalisen median alustoilla käyttäjien sitoutumisen optimoimiseksi. Nämä algoritmit voivat kuitenkin edistää väärän tiedon leviämistä, vaikuttaa julkiseen keskusteluun ja mahdollisesti vahingoittaa yhteisöjä. Utilitarismi kannattaisi suojatoimia, joilla minimoidaan väärän tiedon leviäminen ja säilytetään samalla käyttäjien sitoutuminen alustaan.

- **Yhteiskunnallinen vaikutus:** Autonomisten asejärjestelmien kehittäminen herättää monimutkaisia eettisiä kysymyksiä. Vaikka tekoäly voisi mahdollisesti minimoida ihmisuhrien määrän sodankäynnissä, se herättää myös kysymyksiä vastuullisuudesta ja moraalista, joka liittyy elämää tai kuolemaa koskevien päätösten siirtämiseen koneille. Utilitarismi kehottaisi harkitsemaan huolellisesti yhteiskunnallisia vaikutuksia ja tällaisen teknologian käyttöönoton mahdollisia pitkän aikavälin seurauksia.

Velvollisuusetiikka (Deontologinen etiikka) - toinen vaikutusvaltainen teoria, jossa asetetaan etusijalle tekojen luontainen oikeellisuus tai vääräys riippumatta niiden lopputuloksesta. Tässä lähestymistavassa korostetaan vakiintuneita moraalisia sääntöjä ja periaatteita, kuten ihmisarvoa, käyttäjän suostumusta, yksityisyyden suojaa ja eettisiä normeja. Näiden periaatteiden tulisi ohjata kehittäjiä tekoälyjärjestelmien suunnittelussa ja käyttöönotossa.

- **Yksilöllinen vaikutus:** Se herättää kuitenkin myös huolta käyttäjien suostumuksesta ja algoritmien mahdollisesta vinoumasta. Velvollisuusetiikka edellyttäisi, että kasvojentunnistukseen annetaan selkeä käyttäjän suostumus ja että järjestelmän mahdollisia vinoumia seurataan jatkuvasti.

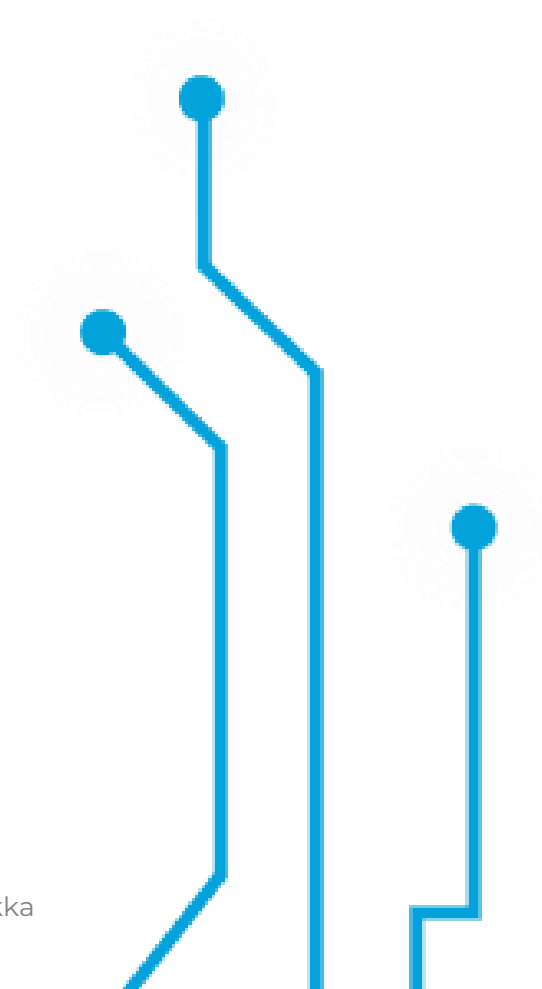


- **Vaikutus yhteisöön:** Ennakoivan poliisitoiminnan algoritmit, joilla pyritään vähentämään rikollisuutta tunnistamalla riskialttiita alueita, herättävät huolta oikeudenmukaisuudesta ja mahdollisesta syrjinnästä. Velvollisuusetiikka puoltaisi sen varmistamista, että nämä algoritmit ovat puolueettomia eivätkä kohdistu suhteettomasti syrjäytyneisiin yhteisöihin.
- **Yhteiskunnallinen vaikutus:** Tekoälyn lisääntyvä käyttö autonomisissa asejärjestelmissä aiheuttaa moraalisen dilemman. Velvollisuusetiikka herättäisi todennäköisesti kysymyksiä siitä, millaisia eettisiä vaikutuksia on, kun elämää tai kuolemaa koskevia päätöksiä annetaan koneiden tehtäväksi, ja siitä, miten tärkeää on säilyttää ihmisen vastuullisuus sodankäynnissä.

Hyve-etiikka - Aristoteleen innoittamana siinä korostetaan moraalisen luonteen ja hyveiden kuten rehellisyys ja empatia vaalimista. Sovelletuna tekoälyn kehittämiseen tämä teoria kannustaa edistämään eettistä käyttäytymistä ja päätöksentekoa alalla. Tämä tarkoittaa sellaisten tekoälyjärjestelmien rakentamista, jotka asettavat etusijalle empatian, avoimuuden ja rehellisyyden kaltaiset piirteet.

- **Yksilöllinen vaikutus:** Sosiaalisen median alustojen kehittäjät asettavat usein käyttäjien sitoutumisen etusijalle tulojen lisäämiseksi. Hyve-etiikkaan perustuva lähestymistapa edellyttäisi kuitenkin, että ne ottaisivat huomioon alustojensa laajemmat yhteiskunnalliset vaikutukset, kuten riippuvuuden tai väärän tiedon leviämisen mahdollisuuden.

- **Vaikutus yhteisöön:** Tekoälyn avulla toimivat työn automatisointiteknologiat voivat lisätä tuottavuutta ja tehokkuutta. Ne herättävät kuitenkin myös huolta työpaikkojen siirtymisestä, mikä johtaa työttömyyteen ja taloudelliseen eriarvoisuuteen. Hyve-etiikan näkökulma kannustaisi kehittäjiä pohtimaan näitä mahdollisia seurauksia ja etsimään keinoja lieventää yhteisöihin kohdistuvia kielteisiä vaikutuksia.
- **Yhteiskunnallinen vaikutus:** Tekoälyä käytetään terveydenhuollossa yhä enemmän diagnosoinnin ja hoitosuunnitelmien parantamiseen. Potilaskeskeinen hoito perustuu kuitenkin vahvasti empatiaan ja myötätuntoon. Hyve-etiikka kannattaisi sen varmistamista, että tekoälyjärjestelmät terveydenhuollossa täydentävät ihmisen arviointikykyä ja asettavat eettiset näkökohdat etusijalle tehokkuuden ja tarkkuuden ohella.

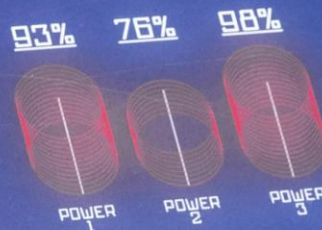


PRIMARY 1

1126674689JF9
9640582
85739
9476
0005830486
8596
9500683
9684
255-OPTIONAL

134134
49923NFSLD95833
97390834
4J4068163-4
5U4793JJFLEO
9375021
595
95821
0586-PORTAL

LIGHT SPEED POWER
901 0985831 9902



POINT DEFENCE
50/10000

HEAT SINK
3/10

SIGNAL SHIP
97%

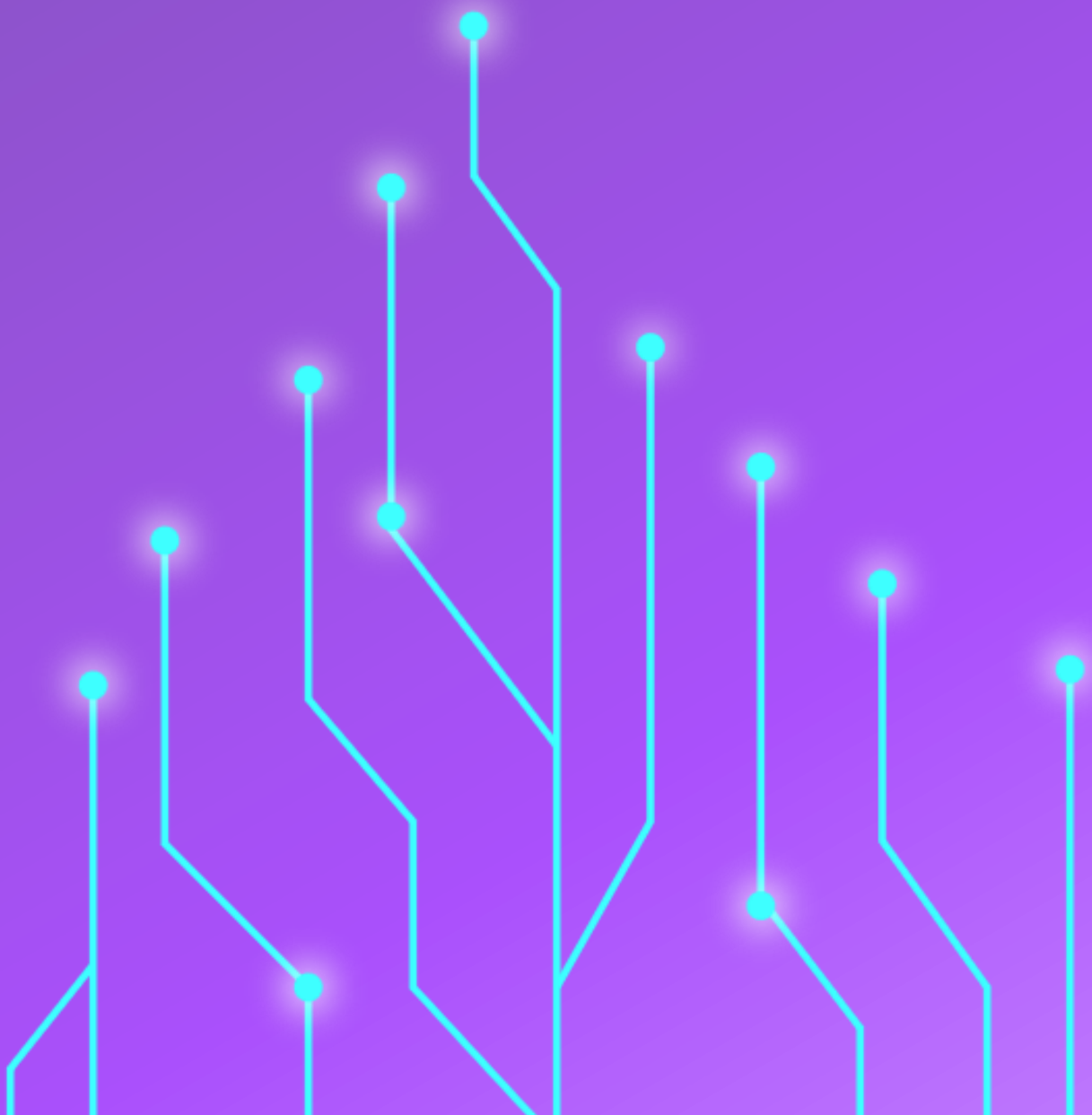
HEAT SINK
1/3

CHAFF
6/10



03. Epäeettisten tekoälykäytäntöjen tunnistaminen ja lieventäminen

Osaamiskokonaisuus 6 | Tekoälyn etiikka käytännössä





03. Epäeettisten tekoälykäytäntöjen tunnistaminen ja lieventäminen

Kuten jo tiedätte, tekoälyn epäeettiset käytännöt voivat ilmetä monissa eri muodoissa, kuten vinoumassa, yksityisyyden loukkauksissa ja avoimuuden puutteessa. Tässä jaksossa käymme nopeasti läpi joitakin tähän mennessä käsiteltyjä tärkeimpiä epäeettisiä käytäntöjä, miten ne voidaan tunnistaa, ja sitten näytämme, miten mahdollinen lieventämisstrategia voisi toimia kunkin osalta.

> Vinoumat ja syrjintä

- **Tunnusmerkit:** Suhteettomat vaikutukset tiettyihin väestöryhmiin. Järjestelmälliset virheet tai erot päätöksenteon tuloksissa. Koulutusdatan monimuotoisuuden puute tai syrjäytyneiden yhteisöjen aliedustus.
- **Esimerkki:** Kasvojentunnistusalgoritmit, joissa värillisten ihmisten virhetaso on korkeampi, mikä johtaa syrjiviin tuloksiin lainvalvonnassa ja tarkkailussa.
- **Ratkaisu:** Monipuoliset datakokonaisuudet ja oikeudenmukaisemmat arviointimittarit

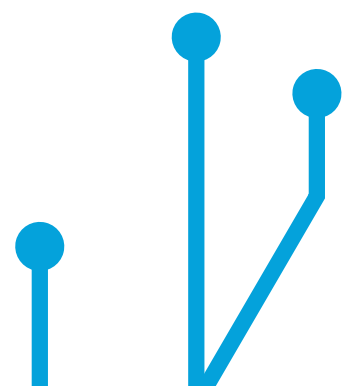


- **Selitys:** Tekoälymallit oppivat datasta, jolla ne on koulutettu. Jos tiedot ovat vinoutuneita, malli ylläpitää näitä vinoutumia. Tämän lieventämiseksi kehittäjät keskittyvät yhä enemmän käyttämään monipuolisia datakokonaisuuksia, jotka edustavat reaali maailman väestöä. Lisäksi kehitetään oikeudenmukaisempia arviointimittareita, joiden avulla voidaan arvioida tekoälymallien vinoumaa ja varmistaa, etteivät ne aseta tiettyjä ryhmiä epäoikeudenmukaisesti epäedulliseen asemaan.



> Läpinäkyvyyden puute

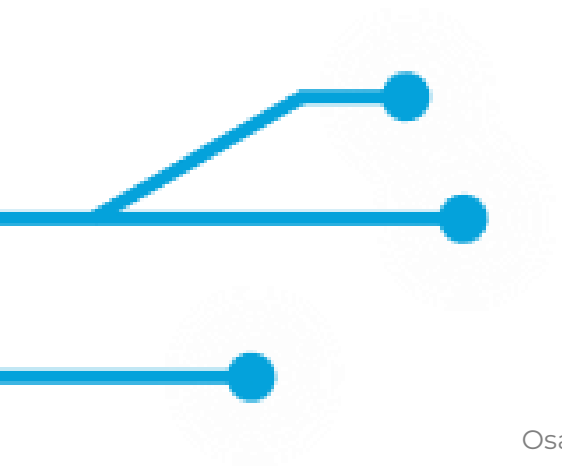
- **Tunnusmerkit:** Mustan laatikon algoritmit, joiden päätöksentekoprosessit ovat vaikeaselkoisia. Tietolähteitä, malliarkkitehtuuria ja päätöksentekoperusteita koskevat tiedot ovat vain rajoitetusti saatavilla. Tekoälyn tuottamien tulosten selitykset tai perustelut puuttuvat.
- **Esimerkki:** Rahoitusalgoritmit, jotka hylkäävät lainahakemuksia antamatta selkeitä syitä tai hylkäysperusteita, mikä lisää epäluottamusta ja turhautumista hakijoiden keskuudessa.
- **Ratkaisu:** Selitettävän tekoälyn (XAI) aloitteet
- **Selitys:** Selitettävissä oleva tekoäly (Explainable AI, XAI) on kasvava ala, joka keskittyy kehittämään tekoälymalleja, jotka ovat läpinäkyvämpiä ja tulkittavampia. Näin ihmiset voivat ymmärtää, miten malli tekee päätöksensä, mikä lisää luottamusta ja mahdollistaa mahdollisten vinoumien tunnistamisen ja niihin puuttumisen. Selitettävä tekoäly voisi esimerkiksi tuoda esiin, mitkä tekijät lainahakemuksen aineistossa vaikuttivat eniten mallin päätökseen.





> Yksityisyyden loukkaukset

- **Tunnusmerkit:** Henkilötietojen luvaton käyttö tai väärinkäyttö. Riittämättömät suojatoimet arkaluonteisten tietojen suojaamiseksi. Tietoon perustuvan suostumuksen saamatta jättäminen tietojen keräämistä ja käsittelyä varten.
- **Esimerkki:** Sosiaalisen median alustat, jotka käyttävät tekoälyalgoritmeja käyttäjien käyttäytymisen ja mieltymysten analysoimiseksi ilman läpinäkyvää tiedonantoa tai kieltäytymisen mekanismeja, mikä vaarantaa käyttäjien yksityisyyden ja autonomian.
- **Ratkaisu:** Tietosuojalainsäädäntö ja käyttäjien valvonta
- **Selitys:** Tietosuoja-asetukset, kuten GDPR (General Data Protection Regulation) ja CCPA (California Consumer Privacy Act), antavat käyttäjille mahdollisuuden hallita henkilökohtaisia tietojaan entistä paremmin. Näiden säännösten mukaan yritysten on oltava avoimia tiedonkeruukäytäntöjen suhteen, hankittava tietoon perustuva suostumus ja annettava käyttäjille mahdollisuus tutustua tietoihinsa, oikaista niitä tai poistaa ne.



> Vastuullisuuden puute

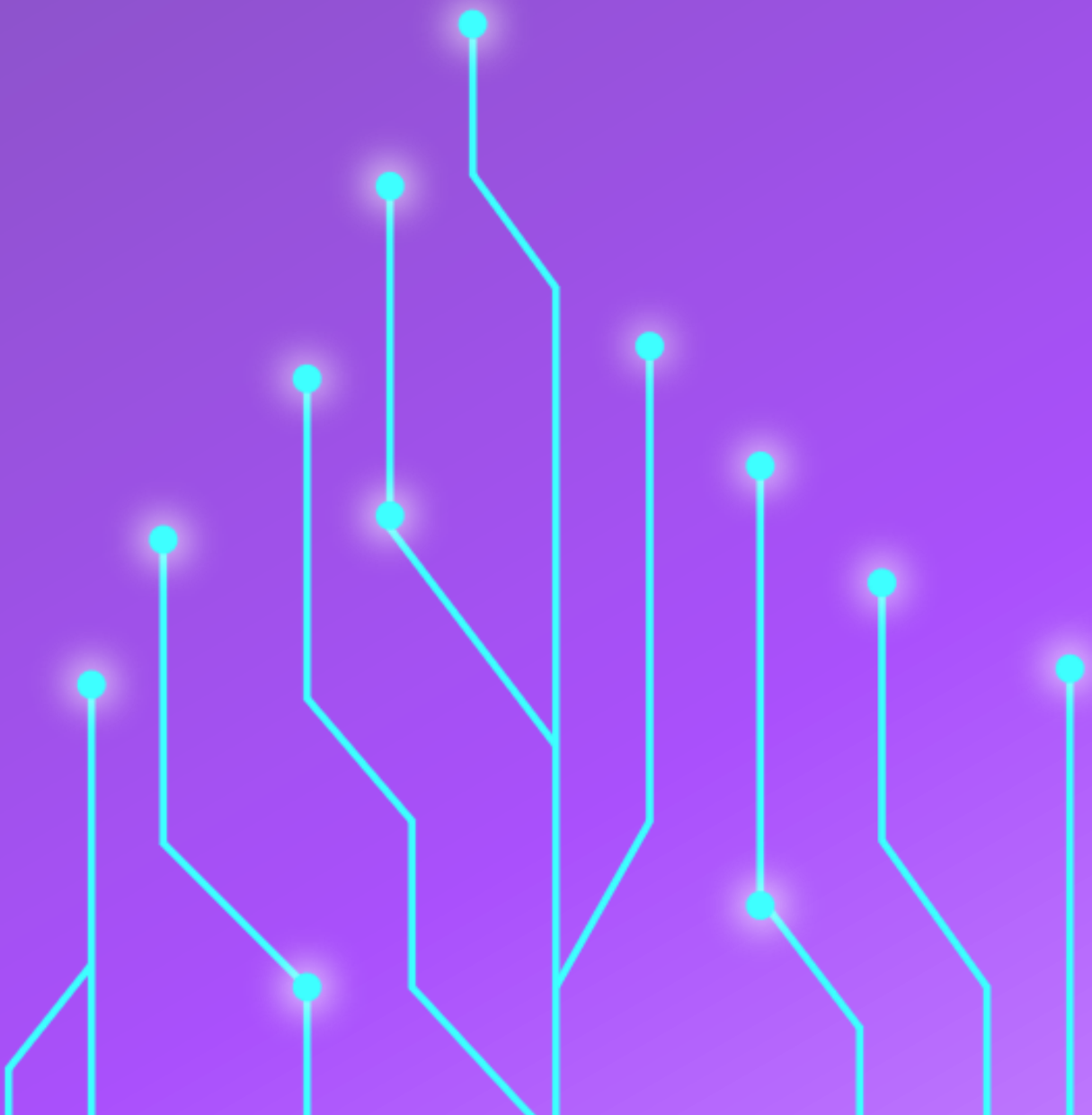
- **Tunnusmerkit:** Mekanismien puuttuminen, joilla vastuulliset osapuolet voitaisiin tunnistaa tai saattaa vastuuseen tekoälyyn liittyvistä haitoista. Rajoitetut oikeussuojakeinot tai oikeussuojakeinot henkilöille, joihin tekoälyn väärinkäytökset vaikuttavat. Tekoälyjärjestelmien vaikutustenarviointien tai tarkastusten tekemättä jättäminen.
- **Esimerkki:** Autonomiset ajoneuvot, jotka ovat joutuneet onnettomuuksiin tekoälyn virheiden vuoksi, jolloin valmistajat välttelevät vastuuta vahingoista, mikä herättää kysymyksiä tekoälyyn perustuvan teknologian oikeudellisesta ja eettisestä vastuusta.
- **Ratkaisu:** Inhimillinen valvonta ja oikeudelliset puitteet
- **Selitys:** Tekoälysovelluksissa, joihin liittyy suuri riski, ihmisen valvonta voi olla ratkaisevan tärkeää. Tämä tarkoittaa sitä, että ihmiset osallistuvat päätöksentekoprosessiin tekoälyn vastuullisen käytön varmistamiseksi ja mahdollisten riskien lieventämiseksi. Lisäksi kehitetään parhaillaan oikeudellisia puitteita, joiden avulla voidaan luoda selkeät vastuuvuorollisuuslinjat tekoälyyn liittyvien haittojen osalta. Näin voidaan varmistaa, että kehittäjät ja yritykset ovat vastuussa tekoälyjärjestelmiensä eettisistä vaikutuksista.



0101

04. Tekoälyn eettisten ohjeiden käytännön soveltaminen

Osaamiskokonaisuus 6 | Tekoälyn etiikka käytännössä



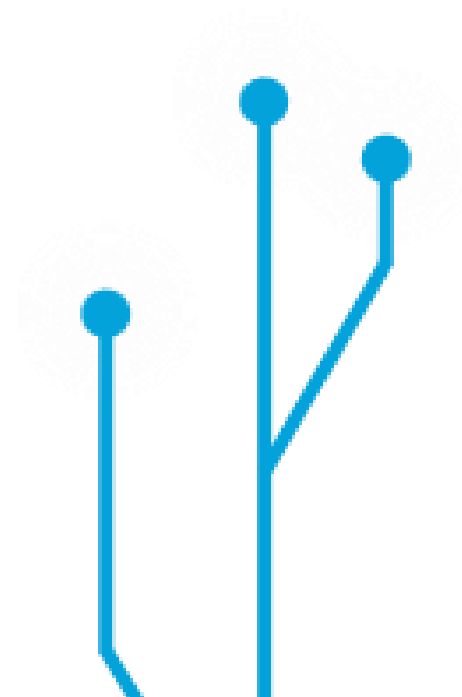


04. Tekoälyn eettisten ohjeiden käytännön soveltaminen

> Eettisen tekoälytyökalupaketin rakentaminen - toimintakehys

Eettiset ohjeet tarjoavat etenemissuunnitelman tekoälyn kehittämiseen ja käyttöönottoon liittyvää eettistä päätöksentekoa varten. Tässä yksikössä olemme käsitelleet eettisten näkökohtien ratkaisevaa roolia tekoälyn kehittämisessä ja käyttöönotossa. Olemme tarkastelleet erilaisia eettisiä koulukuntia, joista jokainen tarjoaa arvokkaat puitteet tekoälyn moniulotteisessa eettisessä maisemassa liikkumiseen. Mutta miten voimme muuttaa nämä abstraktit periaatteet käytännön toiminnaksi?

Tässä osiossa on yksinkertainen nelivaiheinen kehys, jonka avulla voit kehittää omia tekoälyn eettisiä ohjeita.





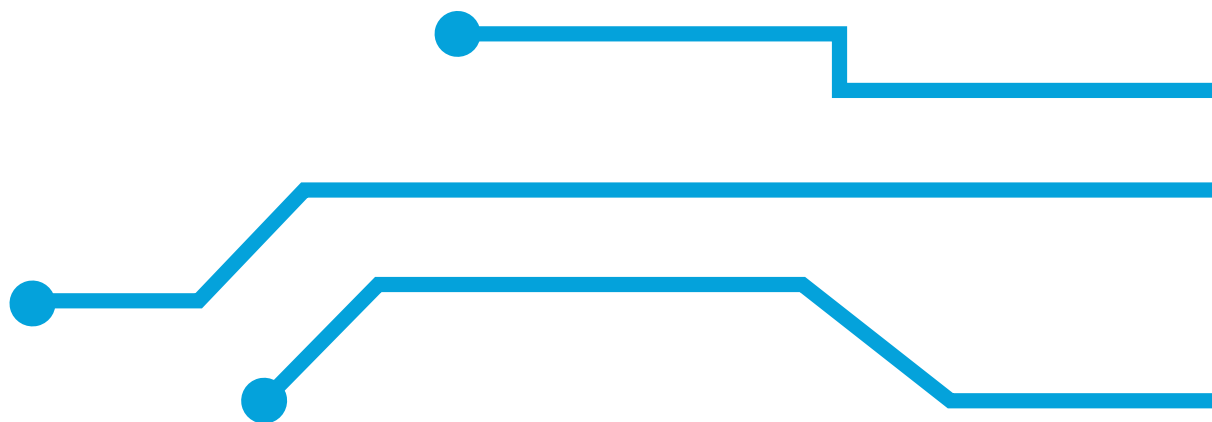
1. Vaihe 1: Tekoälyjärjestelmän tarkoituksen määrittäminen

Ensimmäisessä vaiheessa on määriteltävä selkeästi tekoälyjärjestelmän käyttötarkoitus. Minkä tehtävän suorittamiseen tekoälyjärjestelmä on suunniteltu? Esimerkkeinä voidaan mainita kasvojentunnistusohjelmisto turvallisuustarkoituksiin, tekoälyllä toimiva rekrytointityökalu ansioluetteloiden seulonnan automatisoimiseksi tai lääketieteellinen diagnoosiavustaja terveydenhuollon ammattilaisten tueksi. Järjestelmän tarkoituksen ymmärtäminen on ratkaisevan tärkeää, jotta voidaan tunnistaa asiaan liittyvät sidosryhmät ja eettiset näkökohdat.

2. Vaihe 2: Sidosryhmien ja niiden arvojen tunnistaminen

Kun käyttötarkoitus on määritetty, toisessa vaiheessa on määriteltävä eri sidosryhmät, joihin tekoälyjärjestelmä vaikuttaa. Sidosryhmiin voi kuulua monenlaisia yksilöitä ja ryhmiä, kuten järjestelmän käyttäjät, kehittäjät, ne, joihin järjestelmän päätökset vaikuttavat (esim. tekoälyllä toimivan laina-alustan lainanhakijat), ja yhteiskunta yleensä.

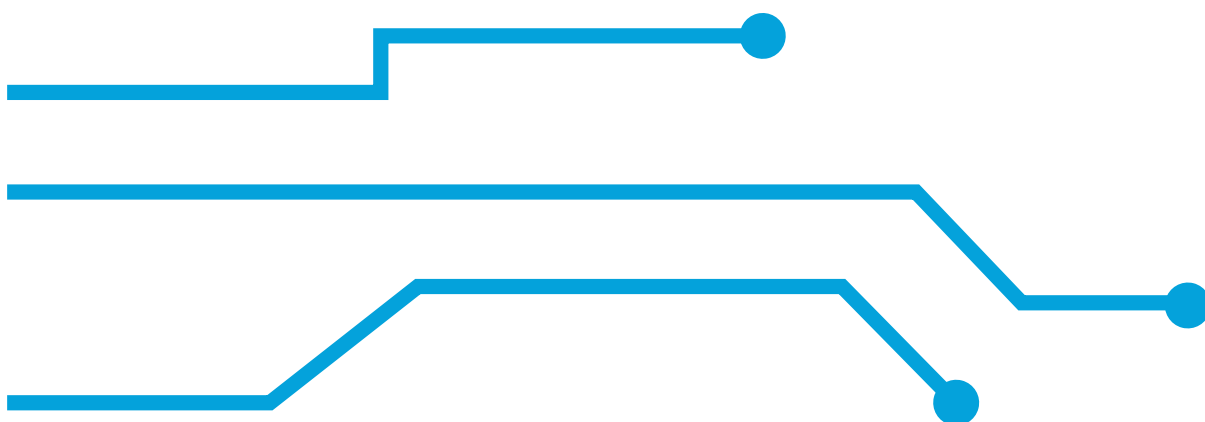
Kullakin sidosryhmällä on omat arvonsa ja prioriteettinsa, jotka on otettava huomioon eettisiä ohjeita laadittaessa. Käyttäjät saattavat esimerkiksi asettaa etusijalle yksityisyyden suojan ja oikeudenmukaisuuden siinä, miten heidän tietojään käsitellään, kun taas kehittäjät saattavat keskittyä järjestelmän toiminnan tehokkuuteen ja tarkkuuteen. On tärkeää tunnustaa nämä erilaiset arvot ja löytää keinoja niiden tasapainottamiseen eettisiä periaatteita muotoiltaessa.





3. Vaihe 3: Keskeisten eettisten periaatteiden esittäminen

Kolmannessa vaiheessa määritellään tekoälyjärjestelmän kehittämistä ja käyttöönottoa ohjaavat keskeiset eettiset periaatteet, jotka perustuvat käyttötapauksen ymmärtämiseen ja tunnistettuihin sidosryhmiin. Nämä periaatteet toimivat perustana tekoälytekniikan vastuullisen ja eettisen käytön varmistamiselle. Joitakin yleisesti harkittuja periaatteita ovat oikeudenmukaisuus, avoimuus, vastuullisuus, yksityisyys ja turvallisuus. Esimerkiksi tekoälyllä toimivan rekrytointityökalun tapauksessa oikeudenmukaisuus edellyttää sen varmistamista, että järjestelmä ei syrji tiettyjä väestöryhmiä vinoutuneiden tietojen perusteella. Avoimuus edellyttäisi selkeää viestintää siitä, miten työkalu toimii ja mitä tekijöitä se ottaa huomioon ehdokkaita arvioidessaan.



4. Vaihe 4: Periaatteiden muuntaminen toteuttamiskelpoisiksi toimenpiteiksi

Viimeisessä vaiheessa abstraktit eettiset periaatteet muunnetaan konkreettisiksi toimiksi, jotka voidaan toteuttaa sen varmistamiseksi, että tekoälyjärjestelmä noudattaa näitä periaatteita. Tähän voi kuulua esimerkiksi erilaisten datakokonaisuuksien käyttäminen, jotta järjestelmän päätöksenteossa esiintyviä vinoumia voidaan lieventää. Avoimuuden lisäämiseksi voidaan käyttää selitettäviä tekoälytekniikoita (Explainable AI, XAI), joiden avulla käyttäjät voivat ymmärtää, miten tekoälyjärjestelmä päätyy tuloksiinsa. Lisäksi on ratkaisevan tärkeää luoda selkeät vastuullisuuslinjat tekoälyjärjestelmän mahdollisesti aiheuttamista haitoista. Tämä voi tarkoittaa vastuun jakamista kehittäjille tai korvausmekanismien luomista niille, joihin tekoälyn päätökset vaikuttavat kielteisesti.

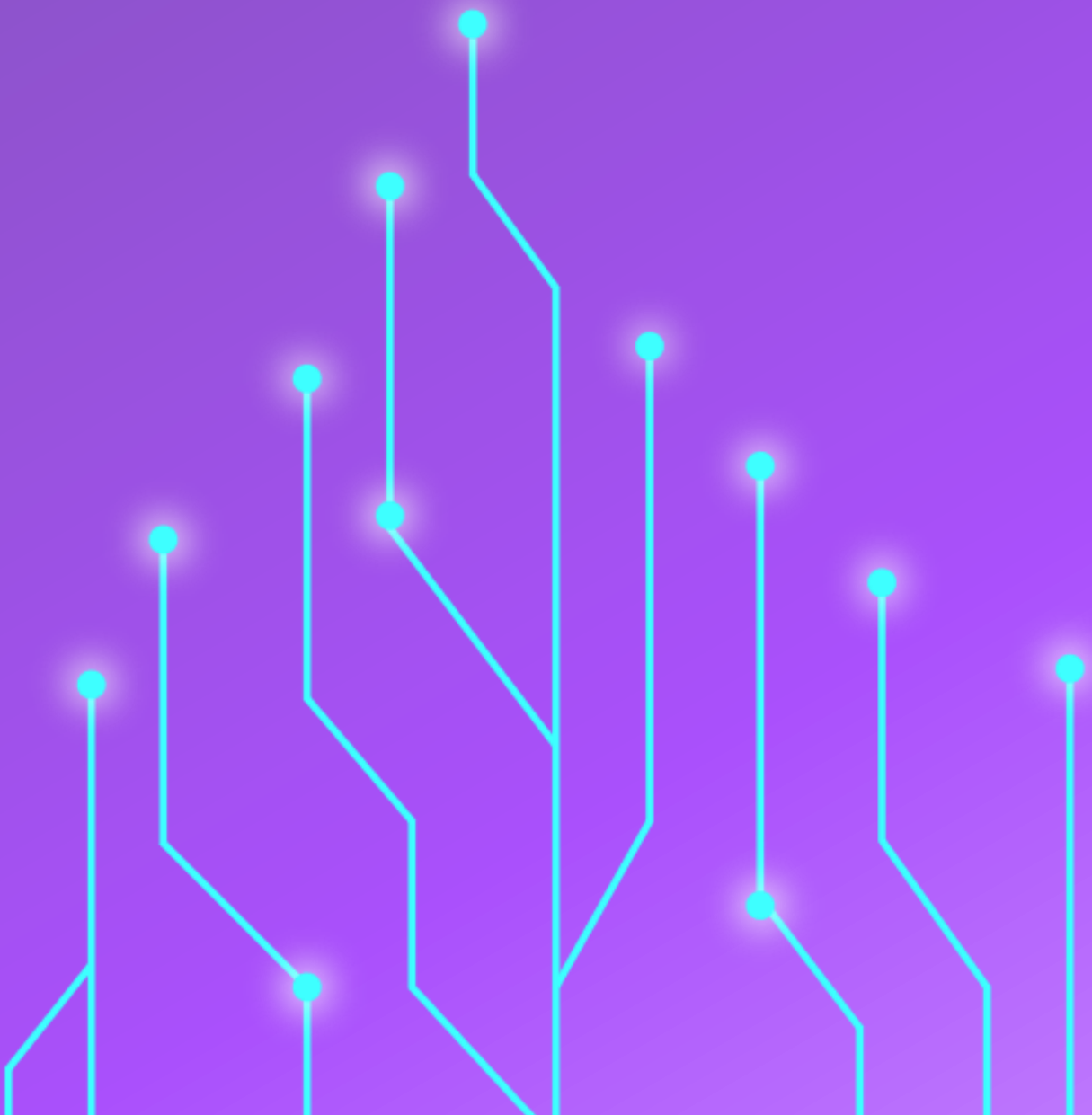
Seuraamalla näitä ohjeita saat valmiudet arvioida tekoälyjärjestelmiä kriittisesti ja edistää näiden tehokkaiden teknologioiden vastuullista kehittämistä ja käyttöä.





05. Vastuullisen tekoälyn kehittämisen ja käyttöönoton edistäminen

Osaamiskokonaisuus 6 | Tekoälyn etiikka käytännössä





05. Vastuullisen tekoälyn kehittämisen ja käyttöönoton edistäminen

> Paremman tulevaisuuden rakentaminen yhdessä - globaali yhteistyö vastuullisen tekoälyn puolesta

Olemme käsitelleet vastuullisen tekoälyn kehityksen merkitystä koko tämän kurssin ajan. Eettisen tekoälyn varmistaminen ei kuitenkaan ole vain paikallinen haaste - se vaatii maailmanlaajuisia ponnisteluja. Tässä tarkastelemme kansainvälisiä aloitteita ja yhteistyöstrategioita, joilla edistetään vastuullista tekoälyn kehittämistä ja käyttöönottoa maailmanlaajuisesti.

- **Tekoälyä koskeva maailmanlaajuinen kumppanuus (GPAI):**

Kyseessä on monen sidosryhmän aloite, johon osallistuu hallituksia, teollisuuden johtajia ja kansalaisyhteiskunnan organisaatioita. GPAI pyrkii kehittämään parhaita käytäntöjä ja suosituksia tekoälyn vastuullista kehittämistä ja käyttöönottoa varten eri aloilla. <https://gpai.ai/>



- **OECD:n tekoälypolitiikan seurantakeskus:**

Tämä Taloudellisen yhteistyön ja kehityksen järjestön (OECD) aloite tarjoaa foorumin, jossa jäsenmaat voivat jakaa tekoälypolitiikan kehittämistä koskevia tietoja ja parhaita käytäntöjä. Sen tavoitteena on edistää kansainvälistä vuoropuhelua ja yhteistyötä tekoälyn eettisestä hallinnoinnista. <https://oecd.ai/>





- **Unescon suositus tekoälyn etiikasta:**

Kyseessä on Unescon vuonna 2021 hyväksymä ei-sitova kansainvälinen instrumentti. Siinä hahmotellaan keskeisiä eettisiä periaatteita tekoälyn kehittämistä ja käyttöä varten ja edistetään ihmisoikeuksia, oikeudenmukaisuutta, läpinäkyvyyttä ja vastuullisuutta.

<https://unesdoc.unesco.org/ark:/48223/pf0000380455#:~:text=AI%20actors%20and%20Member%20States,law%2C%20in%20particular%20Member%20States.>



➤ **EU:n tekoäylaki**

EU:n tekoäylaki on ensimmäinen yritys luoda kattavat puitteet tekoälysovellusten sääntelylle Euroopan unionissa. Euroopan komissio ehdotti sitä ensimmäisen kerran vuonna 2021, ja EU:n parlamentti ja neuvosto hyväksyivät sen lopulta toukokuussa 2024.

Laissa tekoälyjärjestelmät luokitellaan neljään riskiluokkaan - hyväksymiskelvottomaan, korkeaan, rajoitettuun ja vähäiseen - sekä yleiskäyttöistä tekoälyä koskevaan lisäluokkaan. Tekoälysovellukset, joihin liittyy riskejä, joita ei voida hyväksyä, kuten sellaiset, jotka manipuloivat ihmisten käyttäytymistä tai hyödyntävät haavoittuvuuksia, ovat suoraan kiellettyjä.

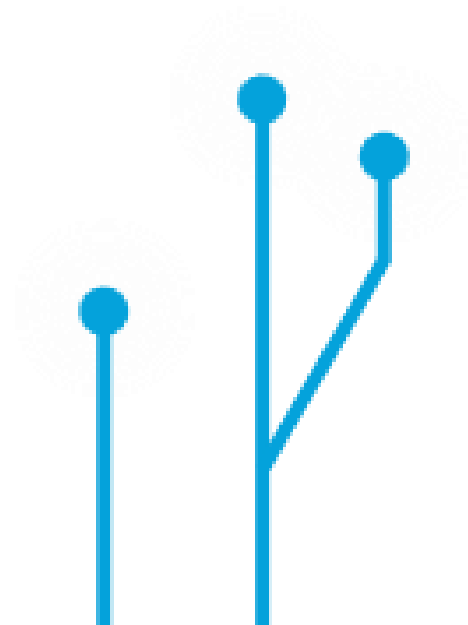
Korkean riskin sovellusten on noudatettava tiukkoja turvallisuus-, avoimuus- ja laatuvaatimuksia, ja niiden vaatimustenmukaisuus on arvioitava vaatimustenmukaisuuden varmistamiseksi. Rajoitetun riskin sovelluksiin sovelletaan avoimuusvelvoitteita, kun taas minimaalisen riskin sovelluksia ei säännellä. Yleiskäyttöisen tekoälyn, erityisesti korkean suorituskyvyn omaavien sovellusten, on täytettävä avoimuusvaatimukset, ja niihin voidaan soveltaa lisäarviointeja.

Tällä riskiperusteisella lähestymistavalla pyritään lieventämään mahdollisia haittoja ja samalla edistämään innovointia ja varmistamaan, että tekoälyn kehittäminen on eurooppalaisten arvojen ja periaatteiden mukaista.

https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf



<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>



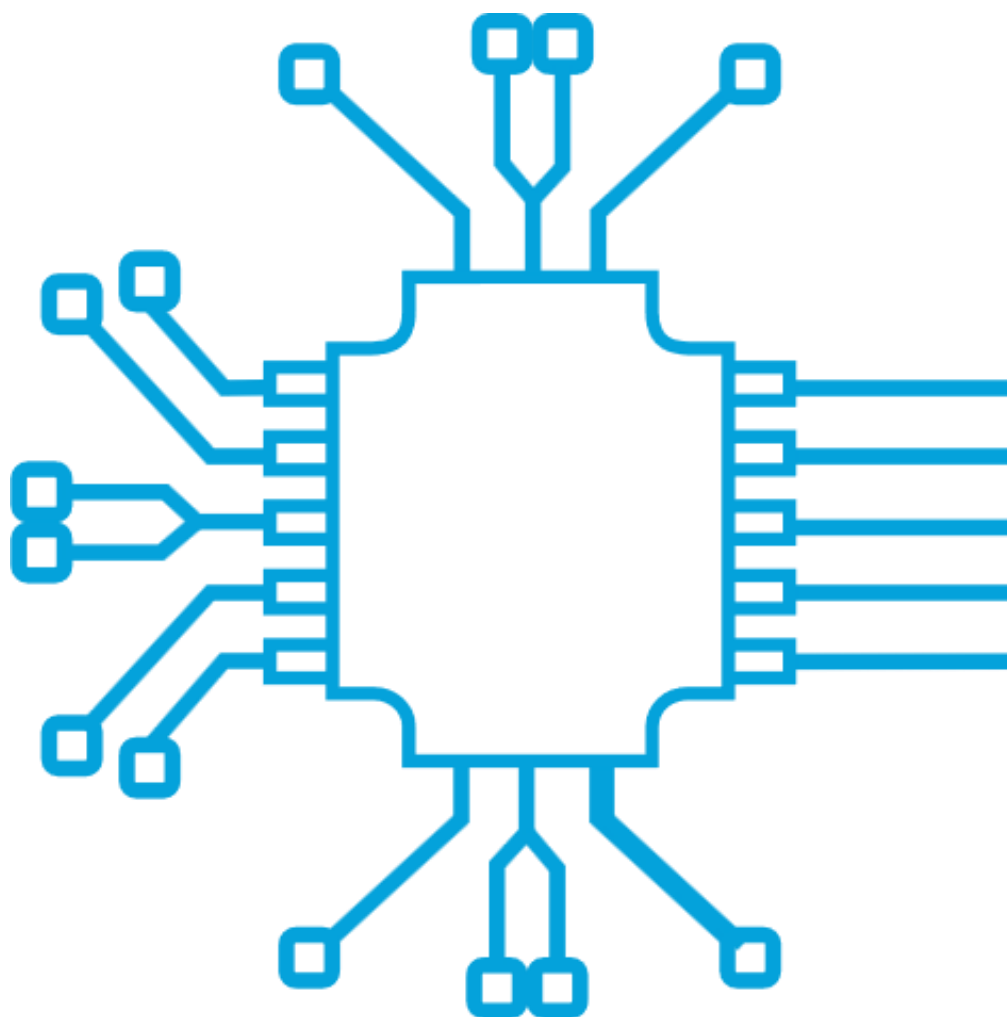


> Sinä VOIT vaikuttaa - Strategioita vastuullisen tekoälytulevaisuuden rakentamiseen

Olet nyt saanut arvokasta tietoa tekoälyn etiikasta ja vastuullisen tekoälyn kehittämisen merkityksestä. Mutta tämä matka ei pääty tähän. Tekoälyn tulevaisuutta muokataan juuri nyt, ja sinun äänelläsi on merkitystä. Seuraavaksi esitellään ehdotuksia käytännön strategioista, joiden avulla voit osallistua ja edistää vastuullisempaa tekoälyn tulevaisuutta omassa yhteisössänne ja muualla.

- **Tietoisuuden lisääminen:** Valitse jokin sinua kiinnostava tekoälyn liittyvä eettinen kysymys (esim. Vinoumat kasvojentunnistuksessa, tekoälyavustajien yksityisyyden suojaan liittyvät ongelmat). Tutki asiaa ja valmistele lyhyt esitys tai työpaja koulun kerhoa, sosiaalikeskusta tai jopa verkkofoorumia varten. Rääätälöi esityksesi yleisölle sopivaksi ja tee siitä mukaansatempaava (käytä visuaalisia ja interaktiivisia elementtejä).
- **Muutoksen puolestapuhuja:** Selvitä olemassa olevat poliittiset aloitteet tai tekoälyn etiikan parissa työskentelevät organisaatiot (esim. paikallishallinnon edustajat, eturyhmät). Tutki heidän työtään ja etsi alat, joilla voit tuoda oman äänesi kuuluviin. Harkitse kirjeen kirjoittamista paikalliselle edustajallesi, jossa ilmaiset huolesi tietystä tekoälysovelluksesta tai puolustat avoimuutta ja vastuuvellvollisuutta koskevia toimenpiteitä.

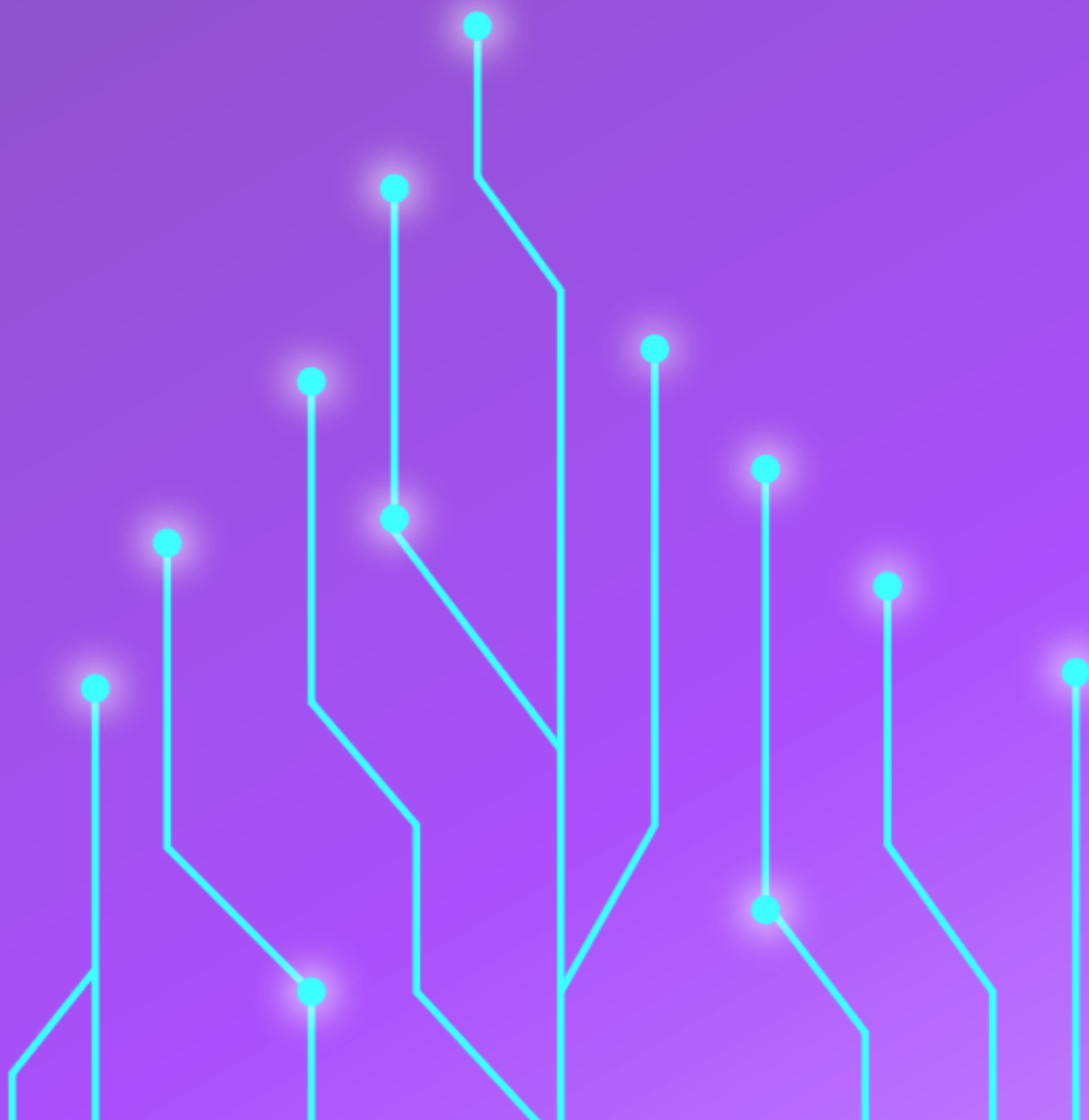
- **Osallistu keskusteluun:** Tutustu tekoälyn etiikkaan keskittyviin verkkofoorumeihin, keskusteluihin ja sosiaalisen median ryhmiin. Osallistu kunnioittaviin keskusteluihin, jaa kurssin aikana oppimasi asiat ja ota oppia muiden näkökulmista. Harkitse oman blogin tai verkkofoorumin perustamista, jotta voit jakaa ajatuksiasi ja lisätä tietoisuutta tekoälyn etiikasta.





05. Yhteenveto

Osaamiskokonaisuus 6 | Tekoälyn etiikka käytännössä





06. Yhteenveto

Tällä kurssilla olemme tutkineet tekoälyn muuttavaa voimaa.

Aloitimme tarkastelemalla algoritmista vinoumaa ja totesimme, että tekoälyn kehittämistä tulisi tarkastella perusteellisesti. 2.

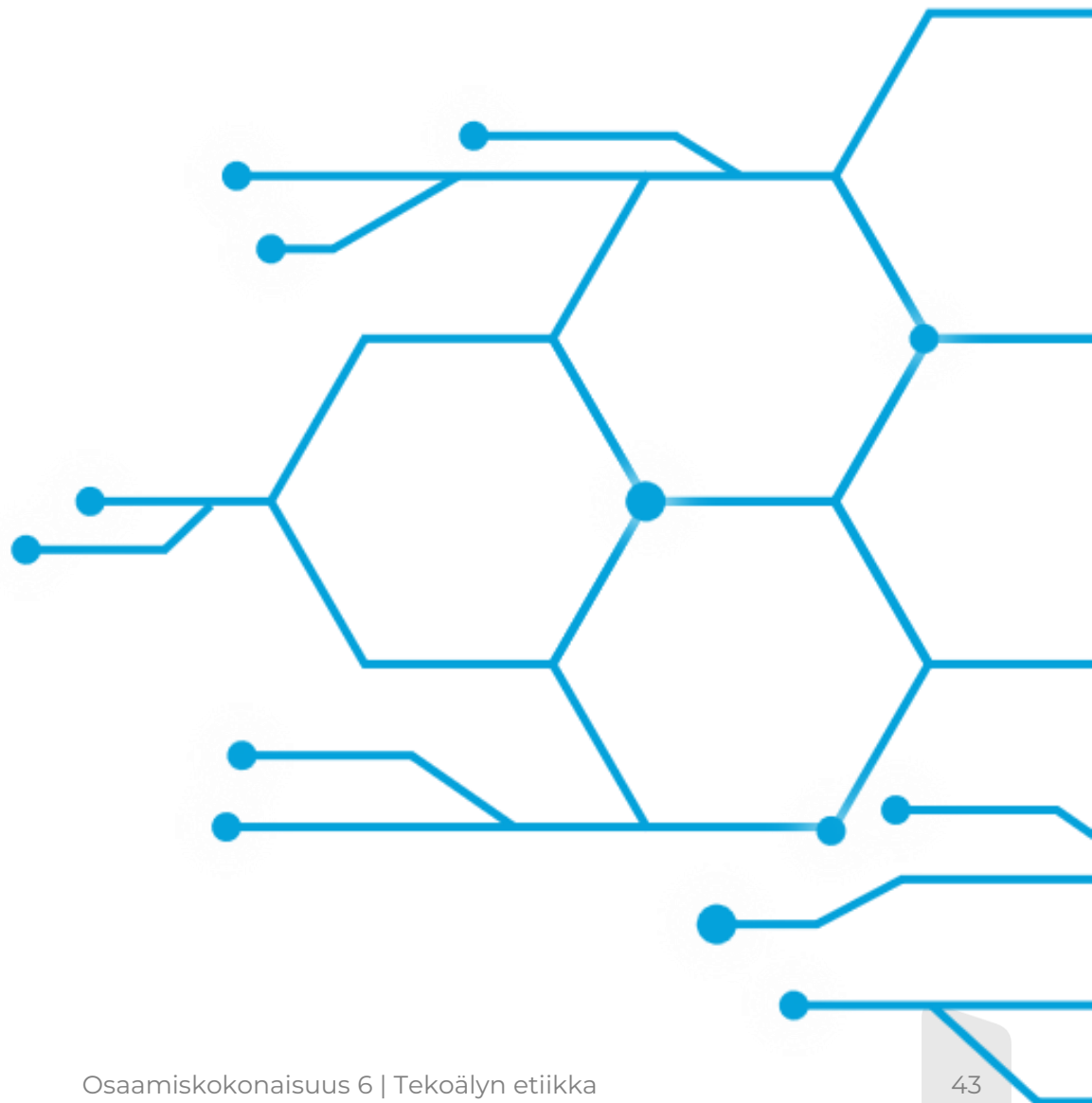
kokonaisuudessa perehdyimme haitan välttäminen -käsitteeseen, jolla varmistetaan, että tekoälyjärjestelmät eivät aiheuta haittaa, ja että ne suunnitellaan turvallisuus edellä. 3. kokonaisuudessa tarkasteltiin vastuullisuuden keskeistä periaatetta ja korostettiin tarvetta selkeisiin vastuualueisiin tekoälyn kehittämisessä ja käyttöönotossa. Kokonaisuudessa 4 keskityttiin läpinäkyvyyteen, jolla varmistetaan ymmärrys siitä, miten tekoälyjärjestelmät toimivat, ja voimme tehdä tietoon perustuvia päätöksiä niiden käytöstä. Kokonaisuudessa 5 käsitelimme tekoälyn ja ihmisoikeuksien välistä yhteyttä ja korostimme oikeudenmukaisuuden ja osallisuuden merkitystä tekoälyn kehittämisessä. 6. kokonaisuudessa annoimme lopuksi käytännön kehityksen omien eettisten tekoälyohjeiden laatimiseen, mikä antaa sinulle mahdollisuuden muuttaa periaatteet toiminnaksi.

Koska olemme tekoälypohjaisen tulevaisuuden kynnyksellä, tällä kurssilla tarkastelemamme eettiset näkökohdat eivät ole pelkkiä harjoituksia, vaan ne ovat vastuullisen tekoälykehityksen perusta.



Osallistumalla aktiivisesti keskusteluun tekoälyn etiikasta, puolustamalla vastuullisia käytäntöjä ja pitämällä kehittäjät ja poliittiset päättäjät vastuullisina voimme varmistaa, että tekoäly palvelee ihmiskuntaa oikeudenmukaisella, tasapuolisella ja hyödyllisellä tavalla.

Tekoälyn tulevaisuutta ei ole ennalta määrätty, vaan se on tulevaisuus, jonka me yhdessä luomme. Rakennetaan siis sellainen tulevaisuus, jossa tekoälystä tulee voima, joka edistää edistystä, hyvinvointia ja valoisampaa huomista kaikille.





Charlie



**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.



**Universitat
de les Illes Balears**



helixconnect



2022-1-ES01-KA220-HED-000085257