



Eettisen tekoälyn mikrokurssi

OPAS

**Osaamiskokonaisuus 1 | Mikä on
algoritminen vinouma?**

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.

VASTUUVAPAUSLAUSEKE: 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. Kurssin sisältö ja odotetut tulokset

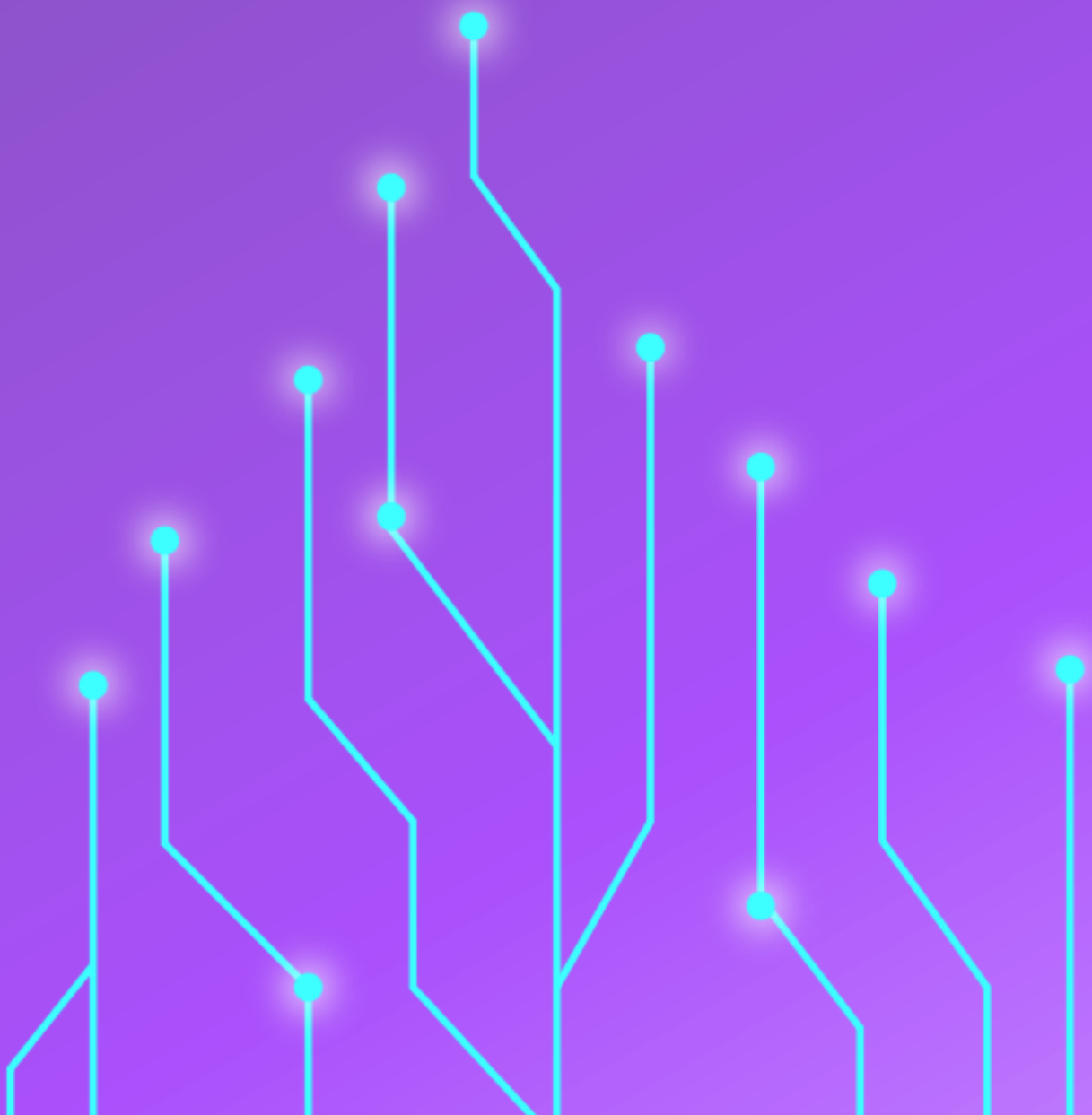
03. Mikä on algoritmisen vinouman?

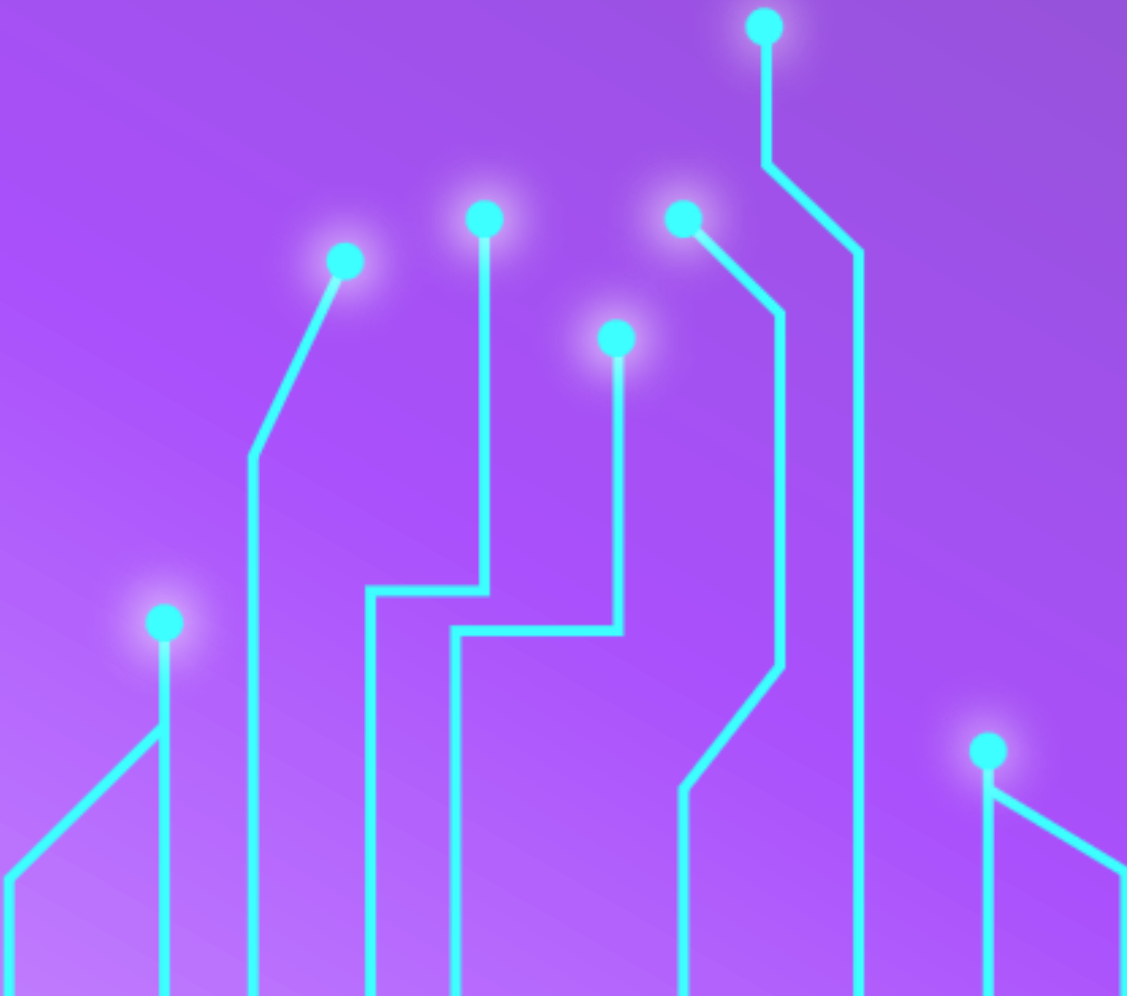
04. Algoritmisen vinouman määritelmä

05. Tekoälyjärjestelmien vinoumien ymmärtäminen

01. Johdanto

Osaamiskokonaisuus 1 | Mikä on algoritminen vinouma?





01. Johdanto

Tekoälyn nopeasti kehittyvässä maisemassa on ensiarvoisen tärkeää varmistaa, että tekoälyteknologian kehittäminen ja käyttö on eettisesti kestävä. Tämä on kattava opas eettisen tekoälyn mikrokurssiin, jossa keskitytään kuuteen osaamiskokonaisuuteen, jotka on suunniteltu antamaan sinulle tiedot ja taidot, joita tarvitset tekoälyn eettisten ongelmien ratkaisemiseen.

Kun lähdet tälle matkalle, tutustut kuuteen eri osaamiskokonaisuuteen, joista jokainen käsittelee tekoälyn eettisen kehittämisen ja käyttöönoton keskeisiä näkökohtia. Nämä osaamiskokonaisuudet on suunniteltu antamaan sinulle tarvittavat työkalut tekoälyteknologioihin liittyvien eettisten haasteiden selvittämiseen algoritmisen vinouman ymmärtämisestä avoimuuden edistämiseen ja ihmisoikeuksien kunnioittamiseen.

Tässä kirjasessa syvennyttään seuraaviin osaamiskokonaisuuksiin:

- OSAAMISKOKONAISUUS 1 - Mikä on algoritmisen vinouma?
- OSAAMISKOKONAISUUS 2 - Haitan välttäminen
- OSAAMISKOKONAISUUS 3 - Vastuullisuus
- OSAAMISKOKONAISUUS 4 - Läpinäkyvyys
- OSAAMISKOKONAISUUS 5 - Ihmisoikeudet ja oikeudenmukaisuus
- OSAAMISKOKONAISUUS 6 - Tekoälyn etiikka käytännössä



Jokaisessa osiossa saat syvällisemmän ymmärryksen tekoälyn keskeisistä eettisistä periaatteista ja käytännöistä sekä käytännönläheisiä näkemyksiä ja käytännön esimerkkejä, jotka vahvistavat oppimaasi.

Olitpa sitten aikuisopiskelija, ammattilainen tai tekoälyn harrastaja, tämä opas on arvokas resurssi, jonka avulla voit laajentaa tietämystäsi ja asiantuntemustasi eettisestä tekoälystä. Kutsumme sinut mukaan tälle matkalle, kun tutkimme tekoälyn eettisiä ulottuvuuksia ja työskentelemme vastuullisemman ja oikeudenmukaisemman tulevaisuuden luomiseksi.

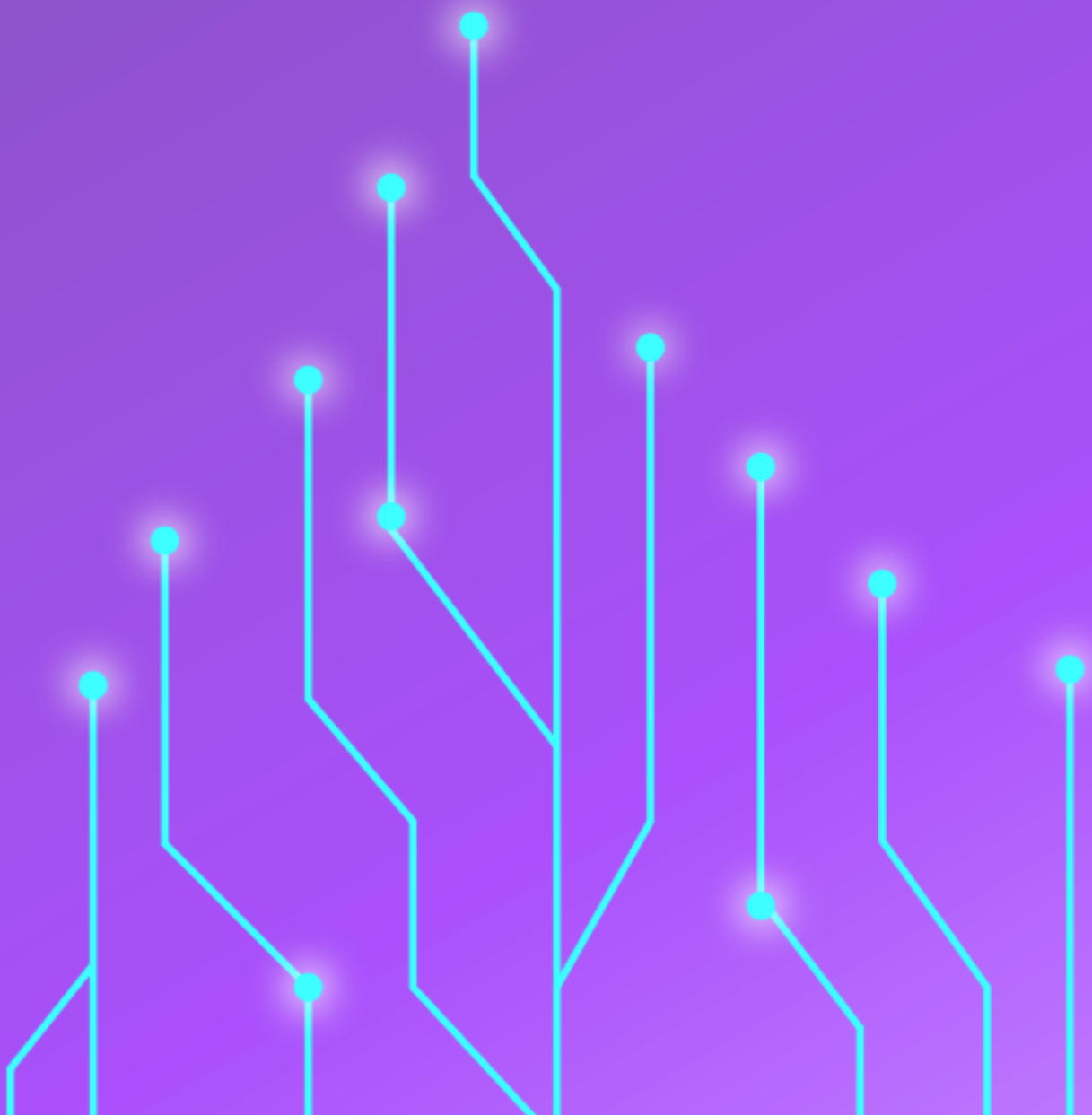
Kiitos, että valitsit tämän oppaan eettisen tekoälyn kehittämiseen ja käyttöön.

Aloitetaan tämä muutoksen tuova matka yhdessä!

CHARLIE-hankeryhmä

02. Kurssin sisältö ja odotetut tulokset

Osaamiskokonaisuus 1 | Mikä on algoritminen vinouma?





02. Kurssin sisältö ja odotetut tulokset

EQF4-tason eettisen tekoälyn mikrokurssi on suunniteltu saavuttamaan seuraavat tulokset:

1. Luodaan perustavanlaatuinen ymmärrys algoritmisesta vinoumasta, tutkia sen alkuperää ja vaikutuksia yksilöihin ja yhteiskuntaan.
 - Tutustutaan algoritmisen vinouman määritelmään, alkuperiin ja ilmenemismuotoihin.
 - Analysoidaan vinoutuneiden algoritmien yhteiskunnallisia ja yksilöllisiä vaikutuksia.
2. Lisätään tietoisuutta siitä, että tekoälyä kehitettäessä on noudatettava eettistä periaatetta, jonka mukaan tekoäly ei saa aiheuttaa haittaa, ja sovelletaan sitä.
 - Arvioidaan vinoutuneisiin algoritmeihin liittyviä riskejä ja haittoja.
 - Kehitetään strategioita haittojen lieventämiseksi ja eettisen tekoälyn kehittämisen edistämiseksi.
3. Arvostetaan vastuun merkitystä tekoälyjärjestelmissä tarkastelemalla asiaa koskevia oikeudellisia ja eettisiä puitteita.
 - Tutkitaan eri sidosryhmien rooleja tekoälyn vastuullisuudessa.
 - Opit parhaat käytännöt vastuullisuuden edistämiseksi tekoälyn kehittämisessä.



4. Tutustutaan **tekoälyjärjestelmien läpinäkyvyyden** käsitteeseen ja sen keskeiseen rooliin algoritmisessa päätöksenteossa.
 - Tutkitaan menetelmiä ja työkaluja tekoälyn avoimuuden lisäämiseksi.
 - Ymmärrät haasteet ja rajoitukset, joita ilmenee kun monimutkaisista algoritmeista tehdään ymmärrettävämpiä.
5. Tutkitaan **tekoälyn, ihmisoikeuksien ja oikeudenmukaisuuden yhtymäkohtia** ja niiden vaikutuksia tekoälyn eettiseen kehittämiseen.
 - Arvioidaan vinoutuneiden algoritmien vaikutusta ihmisoikeuksiin, kuten syrjimättömyyteen, yksityisyyteen ja sananvapauteen.
 - Kehitetään strategioita, joilla varmistetaan oikeudenmukaisuus ja tasapuolisuus tekoälyn kehittämisessä ja käyttöönotossa.
6. Sovelletaan eettisiä periaatteita tekoälyn kehittämisessä ja käyttöönotossa **käytännönläheisten lähestymistapojen ja todellisten skenaarioiden** avulla.
 - Tutkitaan erilaisia eettisiä kehyksiä ja ohjeita sekä niiden soveltamista tekoälyjärjestelmiin.
 - Ymmärrät sidosryhmien sitoutumisen, monitieteisen yhteistyön ja eettisten tekoälyn kehitysprosessien merkityksen.



Kurssin suoritettuaan osallistujat ymmärtävät kokonaisvaltaisesti algoritmiset vinoumat, niiden alakohtaiset vaikutukset sekä välineet ja strategiat niiden käsittelemiseksi. Tämä tieto antaa algoritmipohjaisten alojen ammattilaisille ja tutkijoille/opiskelijoille valmiudet edistää tasapuolisempia ja oikeudenmukaisempia tuloksia datapohjaisessa maailmassa.

Mikrokurssi rakentuu **kuudesta osaamiskokonaisuudesta**, joista jokainen on suunniteltu antamaan osallistujille tiedot ja taidot, joita tarvitaan algoritmisen vinouman haasteiden ja mahdollisuuksien hallintaan.

OSAAMISKOKONAISUUS 1 - Mikä on algoritminen vinouma? Tässä kokonaisuudessa opiskelijat tutustuvat algoritmisen vinouman käsitteeseen ja sen eri ilmenemismuotoihin. Se kattaa sen määritelmän, syyt ja yhteiskunnalliset vaikutukset. Opiskelijat analysoivat vinoumien alkuperää, algoritmeissa olevia lähteitä ja mahdollisia vaikutuksia yksilöihin ja yhteiskuntaan.

OSAAMISKOKONAISUUS 2 - Haitan välttäminen: Tämä kokonaisuus syventyy haitan välttäminen -periaatteeseen, jossa etusijalle asetetaan haittojen välttäminen tekoälyn kehittämisessä ja käyttöönotossa. Osallistujat tutkivat vinoutuneisiin algoritmeihin liittyviä riskejä ja haittoja sekä löytävät strategioita näiden riskien lieventämiseksi ja tekoälyn eettisten käytäntöjen edistämiseksi.



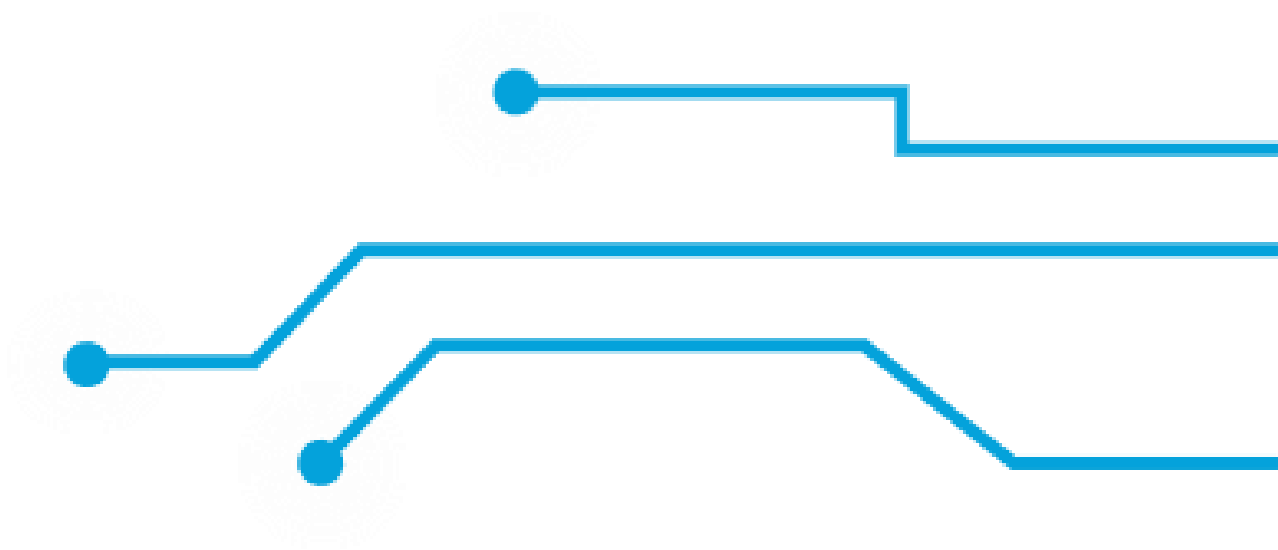
OSAAMISKOKONAISUUS 3 - Vastuullisuus: Tässä kokonaisuudessa opiskelijat syventyvät tekoälyn kehittämisen ja hyödyntämisen vastuullisuuden kriittiseen alueeseen. Osallistujat selvittävät, että on välttämätöntä määritellä selkeät vastuualueet, ja tarkastelevat vastuuvuorollisuutta koskevia oikeudellisia ja eettisiä puitteita. Lisäksi opetussuunnitelmassa tarkastellaan eri sidosryhmien rooleja ja parhaita käytäntöjä, joilla varmistetaan vastuullisuus tekoälyn kehittämistoimissa.

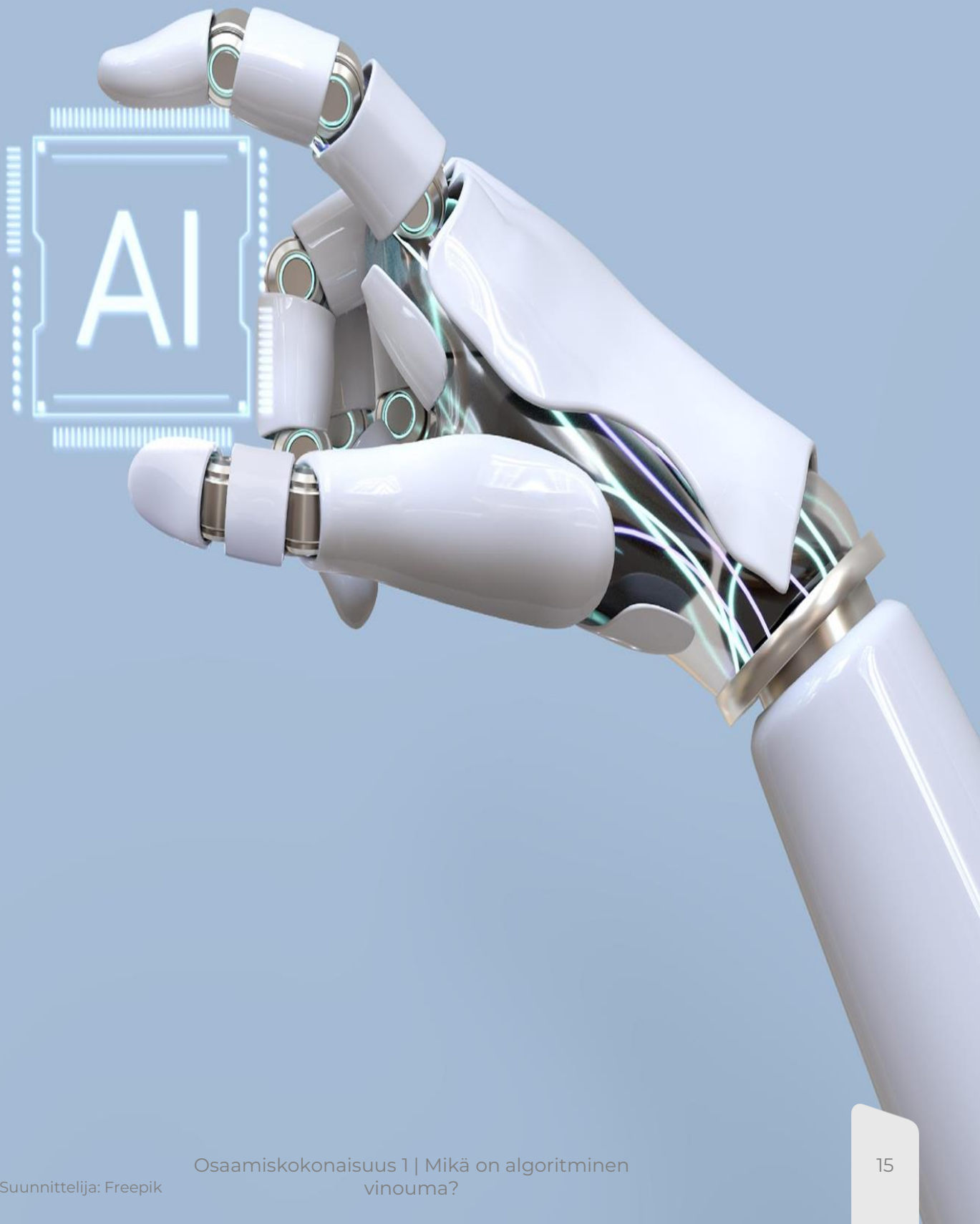
OSAAMISKOKONAISUUS 4 - Läpinäkyvyys: Tämä kokonaisuus valaisee tekoälyjärjestelmien läpinäkyvyyden merkitystä ja korostaa avoimuuden, viestinnän ja selitettävyyden arvoja algoritmisessa päätöksenteossa. Osallistujat tutustuvat tekniikoihin ja resursseihin, joiden tarkoituksena on lisätä tekoälyn avoimuutta, ja samalla he käsittelevät haasteita ja rajoitteita, joita monimutkaisten algoritmien ymmärrettäväksi tekemiseen liittyy.

OSAAMISKOKONAISUUS 5 - Ihmisoikeudet ja oikeudenmukaisuus: Ihmisoikeudet ja oikeudenmukaisuus -kokonaisuudessa opiskelijat tutkivat tekoälyn, ihmisoikeuksien ja oikeudenmukaisuuden yhtymäkohtia. He tutkivat, miten vinoutuneet algoritmit voivat vaikuttaa ihmisoikeuksiin, kuten syrjimättömyyteen, yksityisyyteen ja sananvapauteen. Opiskelijat tutustuvat myös strategioihin, joilla varmistetaan oikeudenmukaisuus ja tasapuolisuus tekoälyn kehittämisessä ja käyttöönotossa.

OSAAMISKOKONAISUUS 6 - Tekoälyn etiikka käytännössä: Tässä kokonaisuudessa korostetaan eettisten periaatteiden käytännönläheistä soveltamista tekoälyn kehittämisessä ja käyttöönotossa. Osallistujat perehtyvät erilaisiin eettisiin kehyksiin ja ohjeisiin ja saavat käsityksen niiden soveltamisesta tekoälyskenaarioissa. Lisäksi kokonaisuudessa korostetaan sidosryhmien sitoutumisen, monitieteisen yhteistyön ja eettisten tekoälyn kehitysprosessien integroinnin merkitystä vastuullisen tekoälyinnovaation edistämisessä.

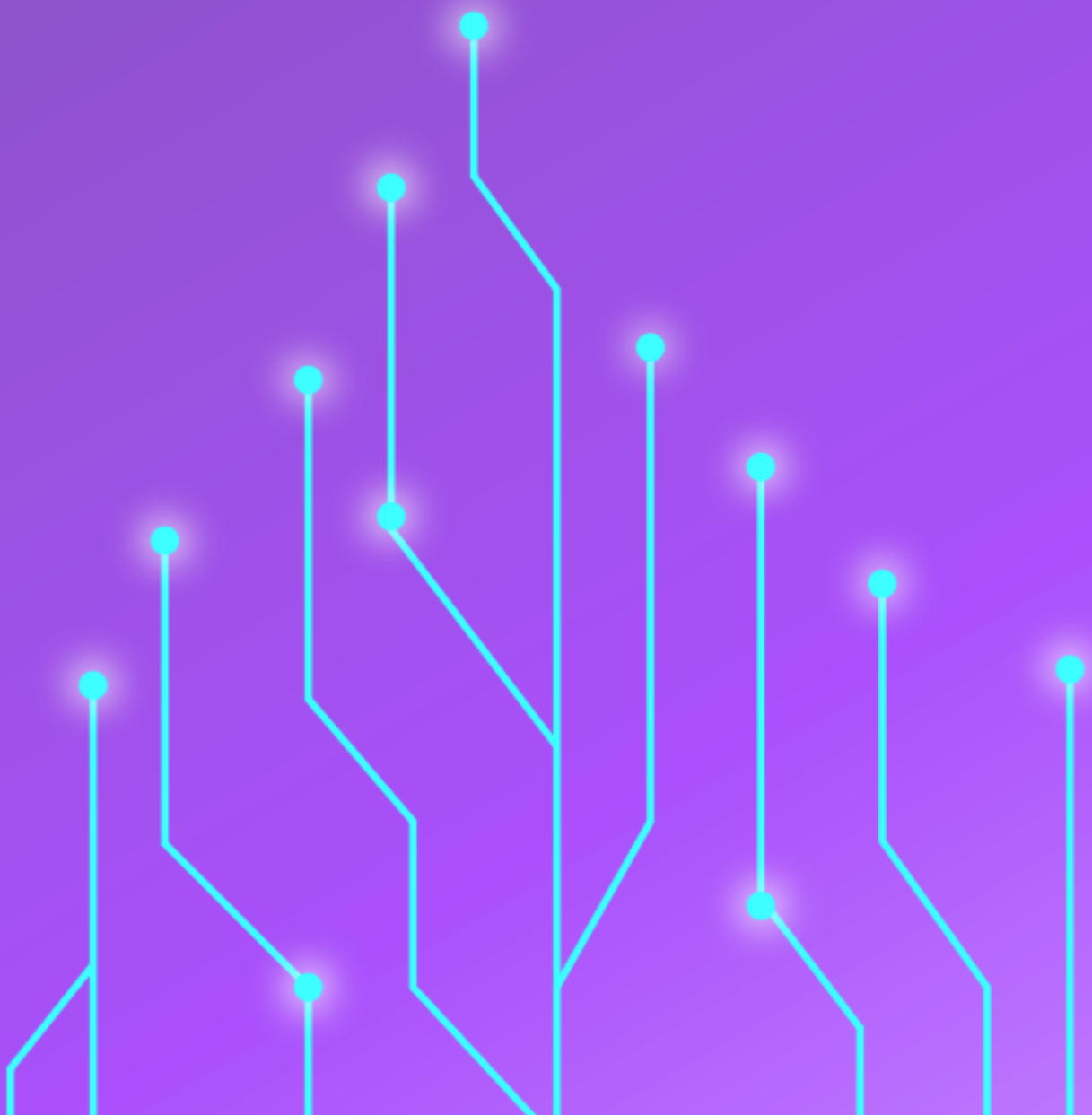
Seuraavassa osiossa käsitellään tarkemmin kunkin osaamiskokonaisuuden sisältöä.





03. Mikä on algoritminen vinouma?

Osaamiskokonaisuus 1 | Mikä on algoritminen vinouma?





03. Mikä on algoritmisen vinouma?

Algoritmeja käytetään tärkeiden päätösten tekemiseen. Ne voivat kuitenkin joskus olla vinoutuneita ja epäoikeudenmukaisia tiettyjä ihmisryhmiä kohtaan. Tätä kutsutaan algoritmiseksi vinoumaksi.

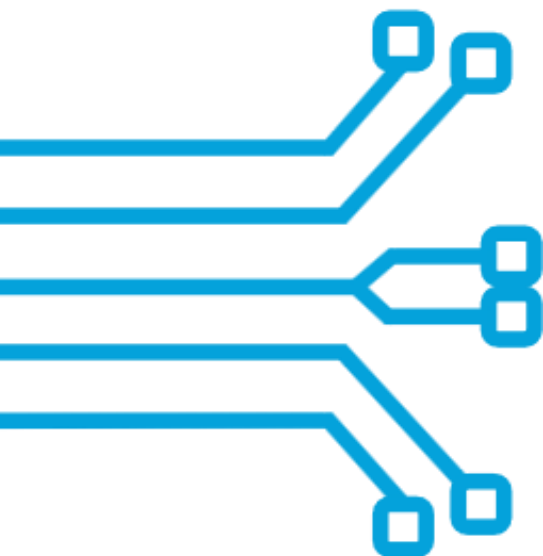
Tässä osaamiskokonaisuudessa opiskelijat oppivat algoritmisesta vinoumasta, sen eri muodoista ja siitä, miten se voidaan tunnistaa. He myös tutkivat algoritmien vinoumien taustalla olevia syitä, mukaan lukien ihmisten vinoumien vaikutus päätöksentekoon. Lisäksi opiskelijat tarkastelevat vinoutuneiden algoritmien mahdollisia seurauksia yksilöille ja yhteiskunnalle, mikä voi johtaa syrjintään ja epäoikeudenmukaiseen kohteluun. Kokonaisuuden lopussa opiskelijat ymmärtävät paremmin algoritmisen vinouman ja sen, miten he voivat käsitellä sitä tulevassa työssään.

Tämän kokonaisuuden osaamistavoitteet ovat seuraavat:

- **Algoritmisen vinouman määrittely:** Opiskelijat oppivat algoritmisesta vinoumasta ja sen syistä, mukaan lukien vinoutunut tiedonkeruu, vinoutunut koulutusdata ja inhimillinen päätöksenteko. Tämä tieto auttaa heitä ymmärtämään, miten vinoumat voivat vaikuttaa tekoälysovelluksiin, kuten kasvojentunnistusjärjestelmiin, jotka tunnistavat väärin tiettyjä ryhmiä.

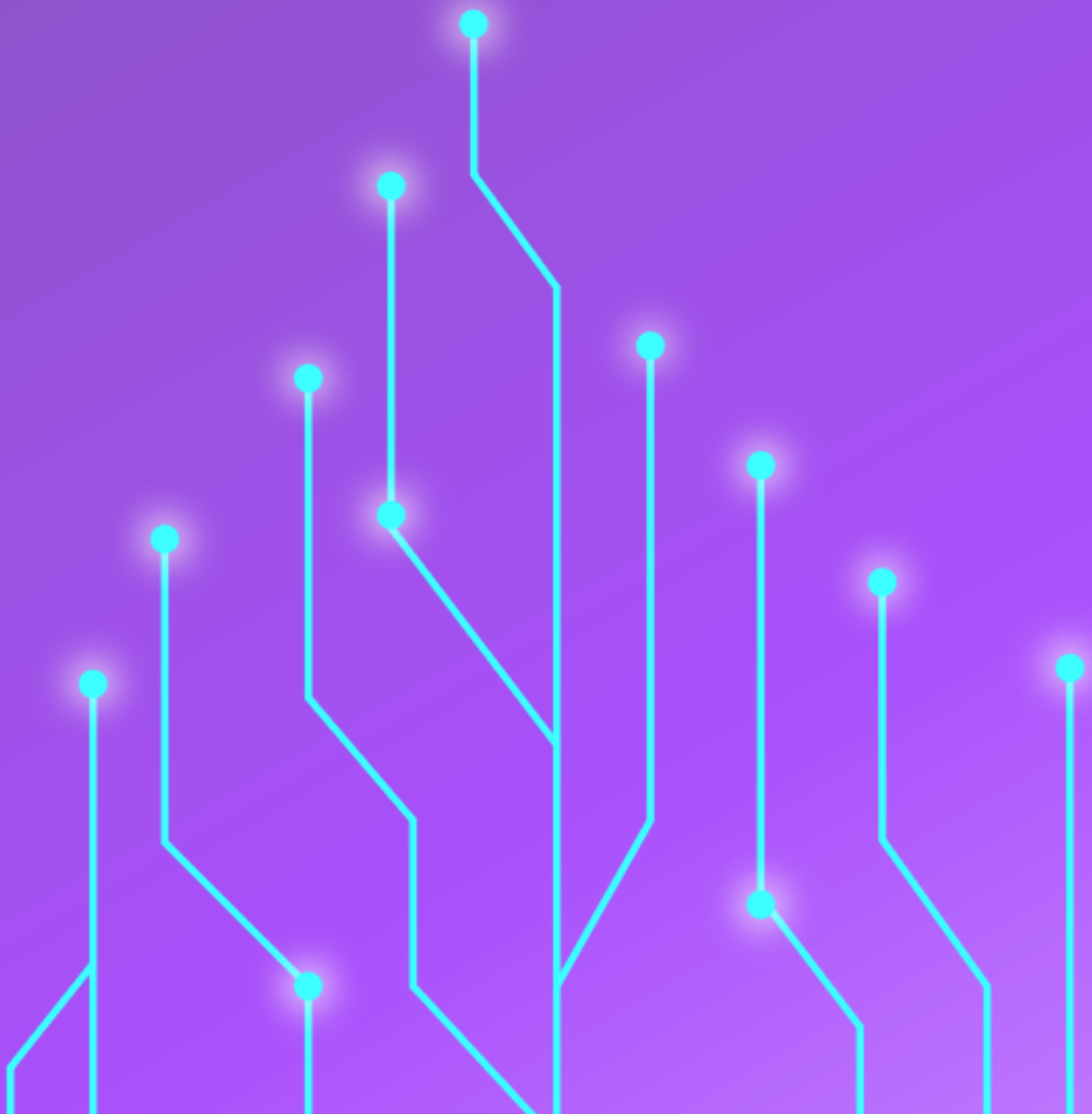


- **Algoritmien vinoumien tunnistaminen:** Opiskelijat oppivat algoritmista vinoumista, mukaan lukien dataan perustuvista, malliin perustuvista ja ihmislähtöisistä ennakkoluuloista. He ymmärtävät, miten nämä vinoumista voivat aiheuttaa epäoikeudenmukaisuutta tekoälyjärjestelmissä. Dataan perustuva vinouma voi esimerkiksi johtua epäedustavasta koulutusdatasta, mikä johtaa vinoutuneisiin ennusteisiin esimerkiksi luottopisteytyksessä tai työnhakijoiden seulonnassa.
- **Algoritmisen vinouman seuraukset todellisessa maailmassa:** Tällä kurssilla opiskelijat tutustuvat algoritmisen vinouman seurauksiin eri aloilla, kuten terveydenhuollossa, rahoituksessa ja rikosoikeudessa. He ymmärtävät, että tekoälyjärjestelmissä on minimoitava algoritmisen vinouman oikeudenmukaisuuden ja tasapuolisuuden edistämiseksi. Kurssilla käsitellään esimerkkejä tekoälyjärjestelmistä, jotka ovat johtaneet kielteisiin tuloksiin terveydenhuollossa ja rikosoikeudessa.



04. Algoritmisen vinouman määritelmä

Osaamiskokonaisuus 1 | Mikä on algoritmisen vinouma?





04. Algoritmisen vinouman määritelmä

Algoritminen vinouma on tekoälyn kriittinen osa-alue, joka on viime vuosina saanut huomiota. Sen ymmärtäminen on olennaista kaikille, jotka osallistuvat tekoälyn kehittämiseen, käyttöönottoon tai sääntelyyn. Määritellään, mitä algoritminen vinouma on ja miksi sen tutkiminen on ratkaisevan tärkeää.

> Mikä on algoritminen vinouma?

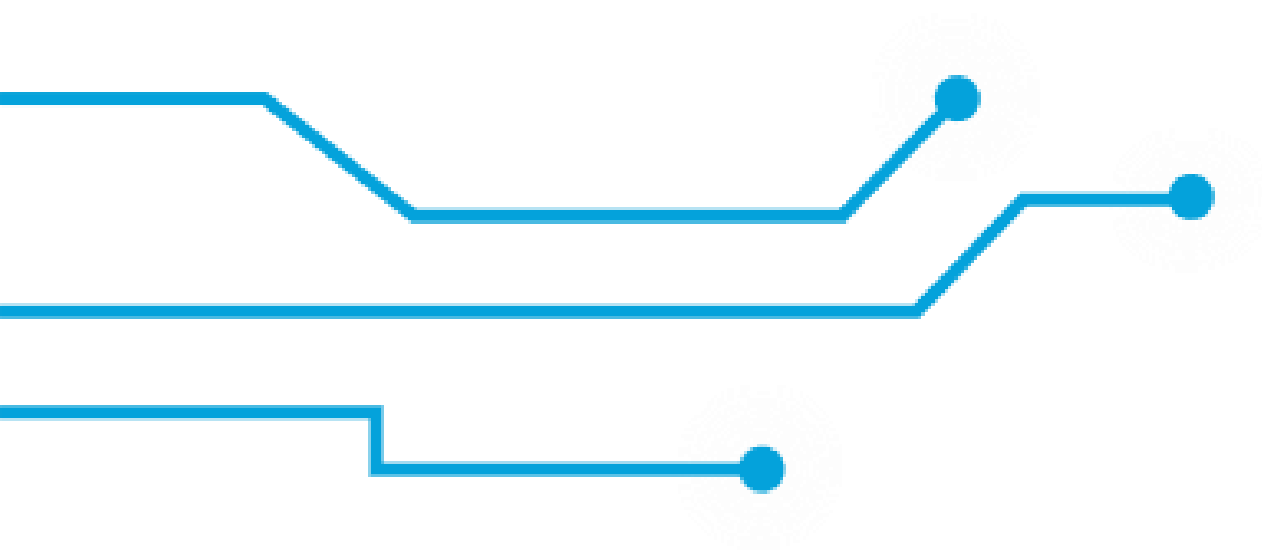
Algoritmisella vinoumalla tarkoitetaan järjestelmällisiä virheitä tai epäoikeudenmukaisuutta tekoälyjärjestelmien tuloksissa, jotka johtuvat erilaisista tekijöistä, kuten vääristyneistä tiedoista, virheellisistä algoritmeista tai inhimillisestä päätöksenteosta. Nämä vinoumat voivat johtaa yksilöiden tai ryhmien syrjivään tai epäoikeudenmukaiseen kohteluun, mikä ylläpitää olemassa olevaa sosiaalista eriarvoisuutta ja vahvistaa stereotypioita.

Miksi tutkia algoritmia?

Algoritmisen vinouman eri muotojen, syiden ja seurausten selvittämiseksi on tärkeää ensin ymmärtää sen määritelmä ja merkitys. Tämän tiedon avulla voimme varustautua työkaluilla, joilla voimme tunnistaa, lieventää ja ehkäistä algoritmista vinoumaa tekoälyjärjestelmissä.



- 1. Eettiset vaikutukset:** Tämä voi johtaa yksilöiden epäoikeudenmukaiseen kohteluun rodun, sukupuolen, iän tai muiden suojattujen ominaisuuksien perusteella, mikä rikkoo oikeudenmukaisuuden ja tasapuolisuuden periaatteita.
- 2. Sosiaalinen vaikutus:** Se voi vaikuttaa syrjäytyneiden yhteisöjen mahdollisuuksiin, resursseihin ja palvelujen saatavuuteen.
- 3. Oikeudelliset ja sääntelyyn liittyvät näkökohdat:** Lainsäätäjät ja sääntelyelimet valvovat yhä tarkemmin algoritmisen vinouman käsittelyä, jotta voidaan varmistaa syrjinnän vastaisten lakien noudattaminen ja suojella yksilöiden oikeuksia.
- 4. Maine ja luottamus:** Tällä voi olla merkittäviä seurauksia niiden brändimielikuvalle ja uskottavuudelle markkinoilla.



➤ **Vinoutuneisiin tuloksiin vaikuttavat tekijät**

Useat toisiinsa liittyvät tekijät edistävät tekoälyjärjestelmien syntymistä, mikä heikentää niiden luotettavuutta, oikeudenmukaisuutta ja tehokkuutta. Tässä jaksossa tarkastelemme joitakin yleisimpiä tekijöitä, jotka vaikuttavat tekoälyjärjestelmien vinoutuneisiin tuloksiin.

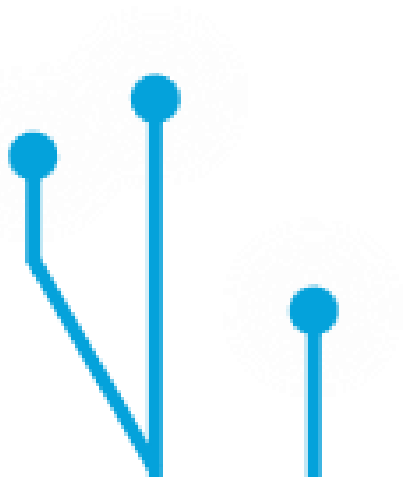
- **Vinoutuneet tiedot:** Tekoälyjärjestelmien kouluttamiseen käytetyt vinoutuneet tiedot johtavat algoritmisiin vinoumiin, jotka voivat johtaa syrjiviin tuloksiin. Tämän lieventämiseksi tietojen keräämiseen ja esikäsittelyyn on kiinnitettävä erityistä huomiota, mukaan lukien edustava otanta, harhojen havaitsemis- ja lieventämisalgoritmit sekä monipuolinen data.
- **Vinoutuneet algoritmit:** Tekoälyjärjestelmät voivat saada vinoutuneita tuloksia virheellisten algoritmien, suunnitteluvalintojen, malliarkkitehtuurien, optimointimenettelyjen tai syötemuuttujien vuoksi. Oikeudenmukaisuustietoinen koneoppiminen, algoritmien läpinäkyvyys ja tulkittavuustekniikat voivat auttaa lieventämään tällaisia vääristymiä.
- **Ihmisen ennakkoluulot:** Tekoälyjärjestelmien vinoumat voivat johtua kehittäjien, tietojenkäsittelytieteilijöiden ja päätöksentekijöiden tiedostamattomista vaikutteista. Näiden vinoumien välttämiseksi tekoälyn kehitystiimien olisi keskityttävä monimuotoisuuteen, eettisiin ohjeisiin ja vastuumekanismeihin.



> Esimerkkejä vinoutuneista järjestelmistä

Tekoälyjärjestelmillä voi olla vinoumia, jotka johtavat epäoikeudenmukaisiin tuloksiin. Seuraavassa on muutamia käytännön esimerkkejä tekoälyjärjestelmistä, jotka ovat yleisesti vinoutuneita ja jotka tuovat esiin algoritmisen vinouman mahdolliset seuraukset. Tutustumme niihin syvällisemmin tämän kurssin myöhemmissä osioissa.

- **Kasvojentunnistusalgoritmit:** Kasvojentunnistustekniikka voi olla vinoutunutta, mikä voi johtaa rotu- tai sukupuolieroihin ja johtaa virheellisiin pidätyksiin tai tiettyjen ryhmien valvontaan. Näiden vinoumien poistaminen on ratkaisevan tärkeää, jotta voidaan varmistaa tekoälyjärjestelmien oikeudenmukaisuus ja tasapuolisuus ja palauttaa yleinen luottamus.
- **Ennakoivan poliisitoiminnan algoritmit:** Ennustavat poliisialgoritmit voivat ylläpitää historiallisissa rikostiedoissa esiintyviä vinoumia, mikä johtaa tiettyjen yhteisöjen tai väestöryhmien liialliseen poliisitoimintaan. Vinoutuneet algoritmit voivat pahentaa lainvalvontakäytäntöjen nykyisiä eroja ja herättää huolta oikeudenmukaisuudesta, vastuuvollisuudesta ja mahdollisista syrjivistä tuloksista rikosoikeusjärjestelmissä..



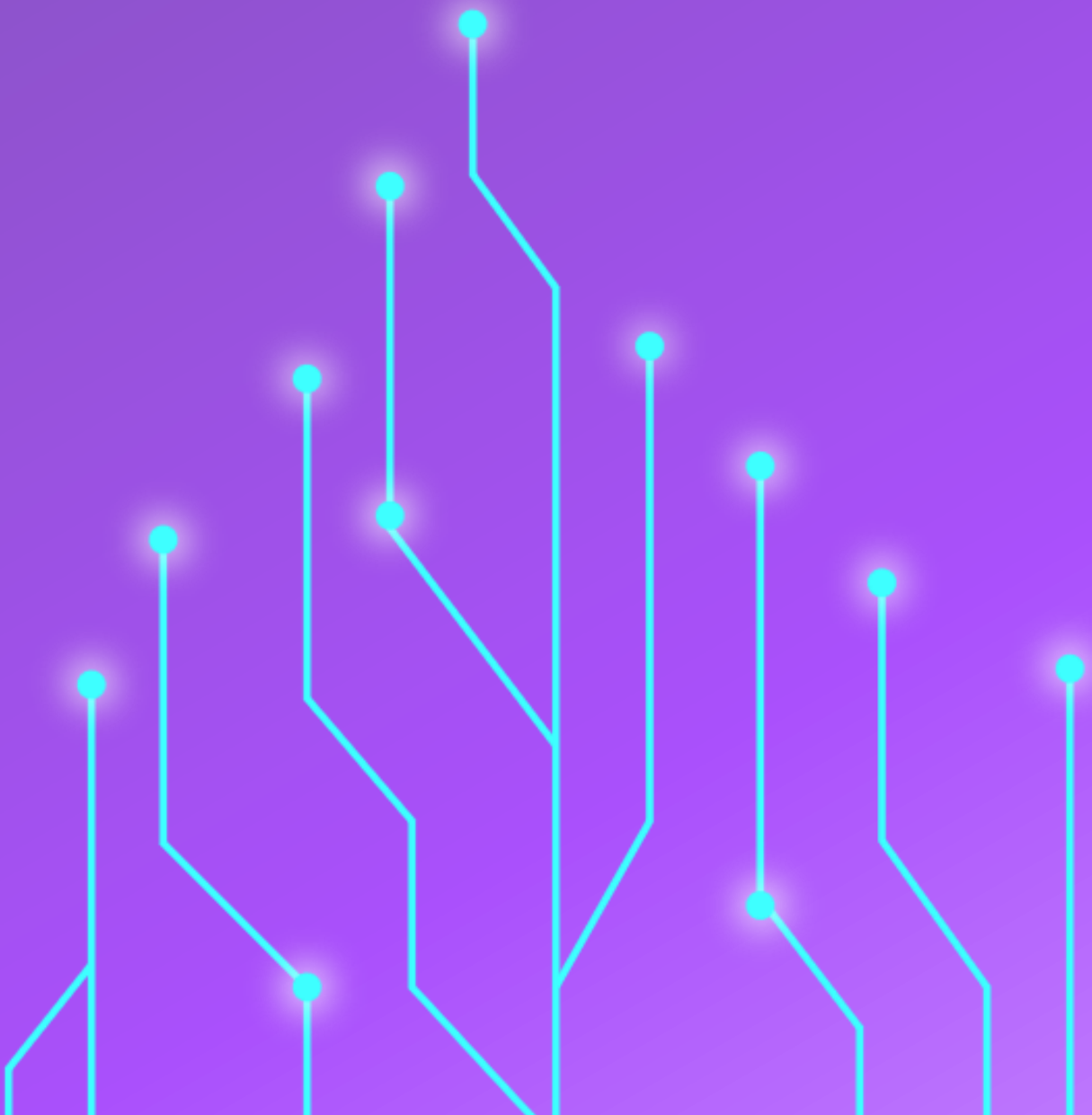
- **Automaattiset rekryointijärjestelmät:** Automaattiset rekryointijärjestelmät voivat ylläpitää ennakkoluuloja, johtaa syrjiviin käytäntöihin ja rajoittaa työvoiman monimuotoisuutta. Vinoutuneet algoritmit voivat oppia vinoumia historiatiedoista, mikä johtaa tiettyjen väestöryhmien suosimiseen. Vinoumien tarkastaminen ja lieventäminen on ratkaisevan tärkeää, jotta voidaan varmistaa oikeudenmukaisuus, tasapuolisuus ja vastuullisuus tekoälyyn perustuvissa rekryointiprosesseissa.



05.

Tekoälyjärjestelmien vinoumien ymmärtäminen

Osaamiskokonaisuus 1 | Mikä on algoritminen vinouma?





05. Tekoälyjärjestelmien vinoumien ymmärtäminen

Tässä jaksossa tarkastelemme kolmea erilaista vinoumaa: **dataan perustuva**, **malliin perustuva** ja **ihmislähtöinen**.

Nämä vinoumat voivat vaikuttaa tekoälyjärjestelmien tarkkuuteen ja luotettavuuteen, ja niiden ymmärtäminen on ensimmäinen askel niiden ehkäisemiseksi.

> **Dataan perustuva vinouma**

Mitä on dataan perustuva vinouma?

Dataan perustuvalla vinoumalla tarkoitetaan vinoumia, jotka johtuvat koneoppimismallien kehittämisessä käytettävän koulutusdatan ominaisuuksista tai jakaumasta.

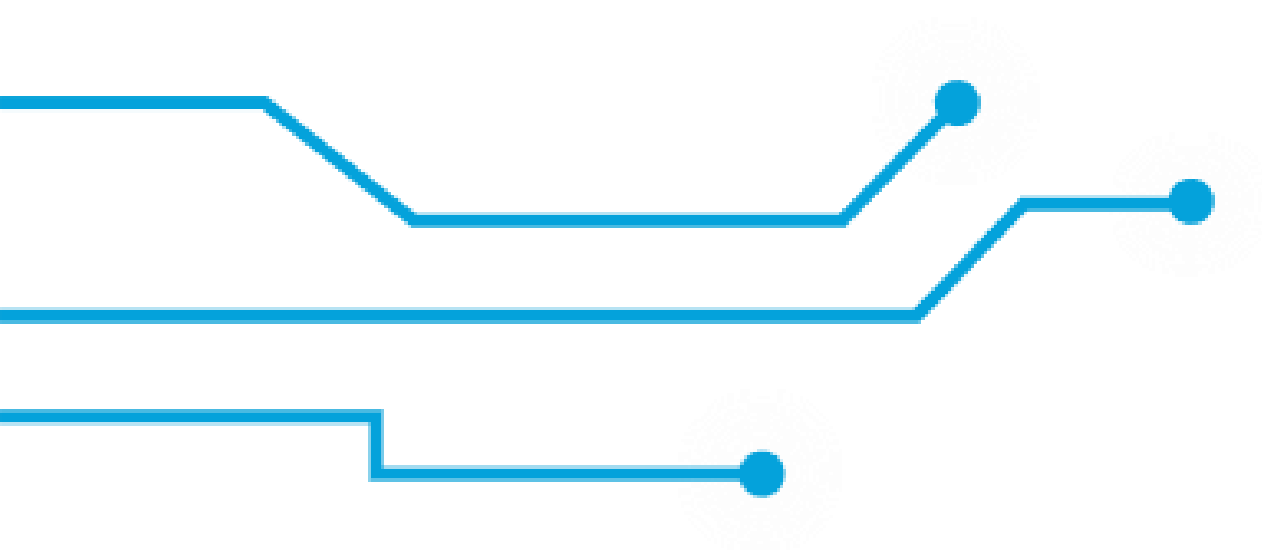
Vinoutunut koulutusdata voi heijastaa historiallista epätasa-arvoa, yhteiskunnallisia ennakkoluuloja tai järjestelmällistä syrjintää, mikä johtaa tiettyjen väestöryhmien vinoutuneeseen edustukseen tai toisten aliedustukseen.

Dataan perustuvan vinouman ymmärtäminen on olennaisen tärkeää, jotta voidaan tunnistaa, miten vinoutunut koulutusdata voi ylläpitää ja pahentaa olemassa olevia stereotypioita, eriarvoisuutta ja syrjiviä käytäntöjä tekoälyjärjestelmissä.



Dataan perustuvien vinoumien syyt

- 1. Puutteellinen tai vääristynyt otanta:** Tämä voi johtaa vinoutuneisiin esityksiin ja vääristyneisiin malliennusteisiin.
- 2. Historialliset ennakkoluulot:** Koulutusdata voi heijastaa yhteiskunnassa esiintyvää historiallista epätasa-arvoa tai järjestelmällisiä vinoumia, jotka ylläpitävät syrjiviä tuloksia tekoälyjärjestelmissä.
- 3. Luokitteluvinoumat:** Vääristyneet tai subjektiiviset luokittelukäytännöt voivat tuoda koulutusdataan vääristymiä, jotka vaikuttavat mallin ennusteisiin ja vahvistavat olemassa olevia stereotypioita.



Esimerkkejä dataan perustuvasta vinoumasta

- 1. Vinoutunut kasvojentunnistus:** Tämä voi johtaa tiettyjen väestöryhmien virheelliseen tunnistamiseen ja syrjintään.
- 2. Sukupuolivinouma kielimalleissa:** Tämä voi heijastaa ja ylläpitää yhteiskunnallisia vinoumia, koska kielimallit, jotka on koulutettu yksipuolisilla tekstikappaleilla, voivat tuottaa sukupuoleen perustuvaa stereotyyppistä tai syrjivää kieltä.
- 3. Rotuun perustuva vinouma ennakoivassa poliisitoiminnassa:** Ennakoivan poliisitoiminnan algoritmit, jotka on koulutettu vinoutuneiden rikostietojen perusteella, voivat kohdistua suhteettomasti vähemmistöyhteisöihin, mikä pahentaa lainvalvonnassa esiintyviä rotujen välisiä eroja.





Dataan perustuvan vinouman vaikutus

- 1. Stereotyyppien vahvistaminen:** Vinoumat voivat vahvistaa olemassa olevia stereotyyppioita ja ennakkoluuloja, mikä ylläpitää syrjintää ja eriarvoisuutta tekoälyjärjestelmissä.
- 2. Eriarvoisuuden vahvistaminen:** Dataan perustuva vinouma voi pahentaa olemassa olevaa eriarvoisuutta ja epätasa-arvoa, mikä johtaa syrjäytyneiden ryhmien epäoikeudenmukaiseen kohteluun ja syrjiviin tuloksiin.
- 3. Luottamuksen heikkeneminen:** Tekoälyjärjestelmät heikentävät luottamusta teknologiaan ja pahentavat huolta oikeudenmukaisuudesta, vastuullisuudesta ja läpinäkyvyydestä.

Dataan perustuva vinouma on merkittävä haaste oikeudenmukaisille ja tasapuolisille tekoälyjärjestelmille. Ymmärtämällä sen syyt ja seuraukset sidosryhmät voivat ryhtyä ennakoiviin toimiin koulutusdatassa esiintyvien vinoumien lieventämiseksi ja tekoälyn osallisuuden edistämiseksi.

> Mallipohjainen vinouma

Mikä on mallipohjainen vinouma?

Mallipohjaisella vinoumalla tarkoitetaan vinoumia, jotka johtuvat koneoppimismallien suunnittelusta, rakenteesta tai optimoinnista ja jotka johtavat syrjiviin tuloksiin tai vinoutuneisiin ennusteisiin.

Mallipohjaisen vinouman syyt

1. **Ominaisuuksien valinnan vinoumat:** Mallinnusprosessin aikana valitut malliominaisuudet voivat tahattomasti koodata koulutusdatassa esiintyviä vinoumia, mikä johtaa vääristyneisiin ennusteisiin tai syrjiviin tuloksiin.
2. **Algoritmien monimutkaisuus:** Monimutkaiset koneoppimisalgoritmit voivat vahvistaa koulutusdatassa esiintyviä vinoumia, mikä vahvistaa niiden vaikutusta mallin ennusteisiin.
3. **Optimointitavoitteet:** Optimointitavoitteet, jotka on määritelty mallin koulutusprosessin aikana, voivat tahattomasti asettaa tietyt tulokset etusijalle muihin nähden, mikä johtaa vinoutuneisiin tai epäoikeudenmukaisiin tuloksiin.





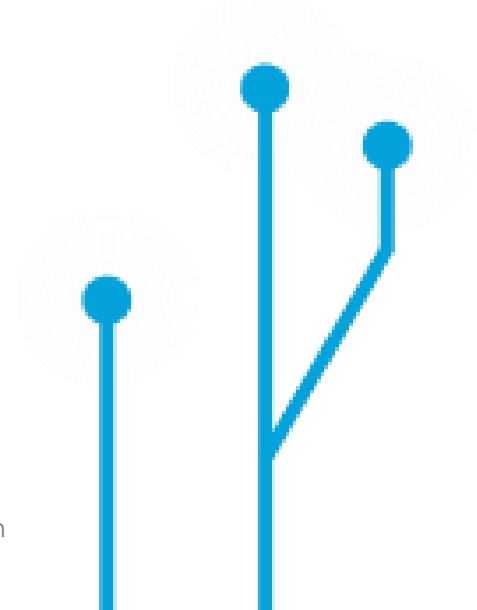
Esimerkkejä mallipohjaisesta vinoumasta

- 1. Sukupuolivinouma rekrytointialgoritmeissa:** Automaattiset rekrytointialgoritmit voivat tahattomasti suosia miespuolisia hakijoita naispuolisiin hakijoihin nähden vinoutuneen ominaisuuksien valinnan tai optimointitavoitteiden vuoksi, mikä ylläpitää sukupuolten välistä epätasa-arvoa rekrytoinnissa.
- 2. Rotuun perustuva vinouma tuomioalgoritmeissa:** Rikosoikeusjärjestelmissä käytettävät ennakoivat tuomioalgoritmit saattavat suositella suhteettomasti ankarampia tuomioita vähemmistöihin kuuluville syytetyille, mikä lisää rotueroja vankeustuomioissa.
- 3. Sosioekonominen vinouma lainojen hyväksymismalleissa:** Lainojen hyväksymisessä käytettävät koneoppimismallit saattavat systemaattisesti evätä lainoja syrjäytyneistä yhteisöistä tulevilta henkilöiltä, mikä pahentaa sosioekonomista eriarvoisuutta rahoituspalvelujen saatavuudessa.

Mallipohjaisen vinouman vaikutus

- 1. Syrjinnän jatkuminen:** Malliin perustuva vinouma voi ylläpitää ja vahvistaa yhteiskunnassa vallitsevaa syrjintää ja eriarvoisuutta, mikä johtaa syrjäytyneiden ryhmien epäoikeudenmukaiseen kohteluun ja vinoutuneisiin tuloksiin.
- 2. Vastuun puute:** Näin ollen tekoälyjärjestelmien syrjivien käytäntöjen tunnistaminen ja niihin puuttuminen voi olla haastavaa.
- 3. Eettiset vaikutukset:** Tämä korostaa tarvetta eettisille ohjeille ja määräyksille, joilla ohjataan tekoälyn kehittämistä ja käyttöönottoa.

Mallipohjainen vinouma asettaa merkittäviä haasteita oikeudenmukaisten ja vastuullisten tekoälyjärjestelmien kehittämiselle ja käyttöönotolle. Ymmärtämällä mallipohjaisen vinouman mekanismeja ja vaikutuksia sidosryhmät voivat toteuttaa strategioita vinouman lieventämiseksi ja tekoälyteknologioiden oikeudenmukaisuuden ja tasapuolisuuden edistämiseksi.





> Ihmislähtöinen vinouma

Mikä on ihmislähtöinen vinouma?

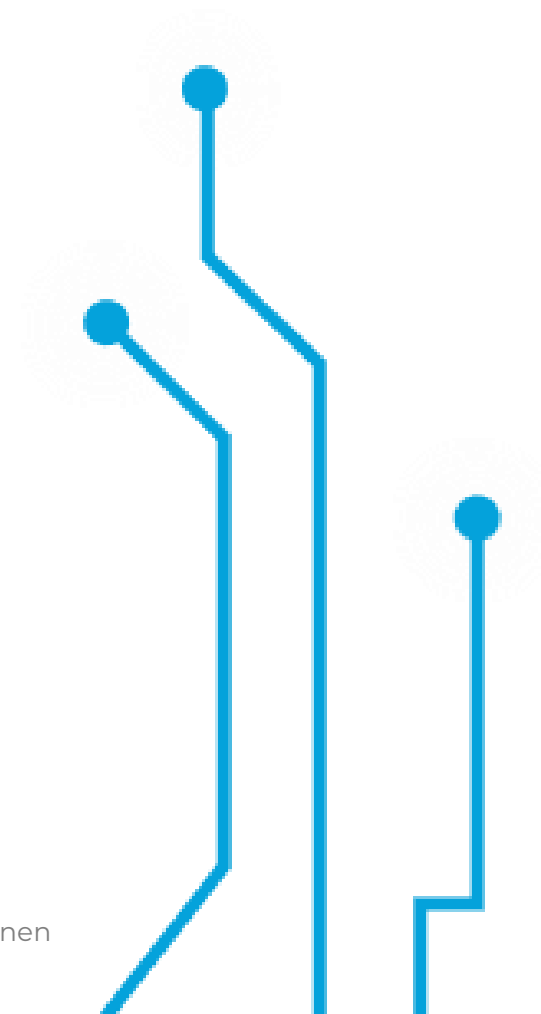
Tekoälyn inhimillisellä vinoumalla tarkoitetaan vinoumia, jotka johtuvat kehitykseen ja käyttöönottoon osallistuvien henkilöiden päätöksistä, toimista tai arvostelusta. Se voi johtua kognitiivisista ennakkoluuloista, kulttuurisista vaikutteista ja yhteiskunnallisista ennakkoluuloista, jotka johtavat vinoutuneisiin tuloksiin tai syrjiviin käytäntöihin.

Ihmislähtöisten vinoumien syyt

- 1. Tiedonkeruun vinoumat:** Tiedonkeruun vinoumat, kuten otanta- tai valintaharhat, voivat johtaa vinoutuneeseen koulutusdataan ja vinoutuneisiin malliennusteisiin.
- 2. Algoritmien suunnittelun vinoumat:** Tekoälyalgoritmit voivat olla vääristyneitä ihmisen suunnittelijoiden ja kehittäjien valintojen vuoksi, jolloin tekoälyjärjestelmissä syntyy vinoutuneita tuloksia.
- 3. Tulkinta- ja käyttöönottovinoumat:** Ihmistulkit ja päätöksentekijät voivat olla tekoälyjärjestelmiä käyttäessään puolueellisia, mikä voi johtaa syrjiviin käytäntöihin ja epäoikeudenmukaiseen kohteluun.

Esimerkkejä ihmislähtöisistä vinoumista

- 1. Kasvojentunnistusjärjestelmien vinoumat:** Tämä voi johtaa tiettyjen väestöryhmien virheelliseen tunnistamiseen tai aliedustukseen. Tämä voi johtaa siihen, että kasvojentunnistusjärjestelmät tunnistavat tiettyjä väestöryhmiä väärin tai ovat aliedustettuina.
- 2. Rekrytointialgoritmien vinouma:** Vinoumat inhimillisissä päätöksentekoprosesseissa, kuten ansioluettelon seulonnassa tai haastattelujen arvioinnissa, voivat ylläpitää sukupuolten tai rotujen välisiä eroja rekrytoinnissa, vaikka käytettäisiin tekoälyyn perustuvia rekrytointialgoritmeja.





Ihmislähtöisten vinoumien vaikutus

- 1. Nykyisen eriarvoisuuden kärjistyminen:** Tekoälyn ihmislähtöiset vinoumat voivat pahentaa yhteiskunnassa vallitsevaa eriarvoisuutta ja epätasa-arvoa. Vinoutunut tiedonkeruu, algoritmien suunnittelu ja tulkinta voivat johtaa syrjäytyneiden ryhmien epäoikeudenmukaiseen kohteluun, mikä ylläpitää syrjintää ja estää yhteiskunnallista edistystä.
- 2. Luottamuksen ja kansalaisten luottamuksen heikkeneminen:** Tekoälyjärjestelmät, joihin liittyy inhimillisiä vinoumia, voivat heikentää yleisön luottamusta teknologiaan. Oikeudenmukaisuuteen, läpinäkyvyyteen ja velvollisuuteen liittyvät huolenaiheet voivat herättää huolta, mikä estää tekoälyn käyttöönottoa ja hyväksyntää eri aloilla.
- 3. Tekoälyjärjestelmien tehokkuuden väheneminen:** Ihmisen aiheuttamat ennakkoluulot voivat heikentää tekoälyjärjestelmien tehokkuutta. Vinoutunut koulutusdata tai ihmisten tekemät vääristyneet tulkinnat voivat johtaa epätarkkoihin ennusteisiin, virheellisiin suosituksiin ja epäoptimaalisiin lopputuloksiin, mikä haittaa tekoälyn mahdollisia hyötyjä.

Ihmisten puolueellisuus on merkittävä haaste oikeudenmukaisten ja vastuullisten tekoälyjärjestelmien luomiselle. Ymmärtämällä ja lieventämällä algoritmisia vinoumia sidosryhmät voivat rakentaa luotettavampia ja läpinäkyvämpiä tekoälyjärjestelmiä.



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