



Eettisen tekoälyn mikrokurssi

OPAS

Osaamiskokonaisuus 3 | Vastuullisuus

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. Vastuullisuuden ymmärtäminen ja vastuullisuuden vaaliminen tekoälyssä

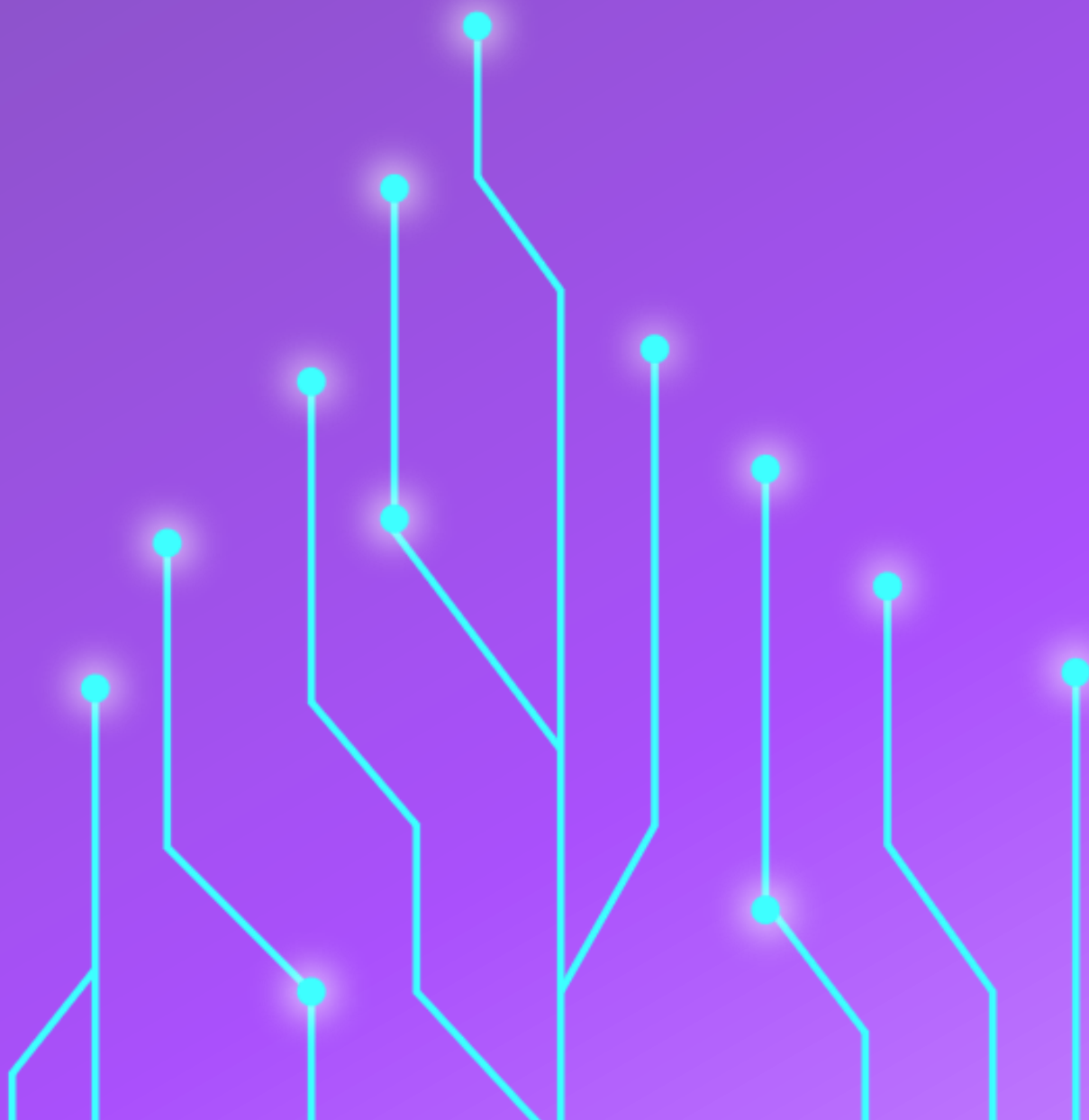
03. Vastuullisuus: Algoritmiseen vinoumaan puuttuminen

04. Vastuumekanismit, joilla puututaan algoritmiseen vinoumaan

05. Yhteenveto

01. Johdanto

Osaamiskokonaisuus 3 | Vastuullisuus





01. Johdanto

Tässä osaamiskokonaisuudessa oppijat hankkivat perustiedot tekoälyn vastuullisuudesta keskittyen peruskäsitteiden ymmärtämiseen, tekoälyn kehittäjien ja käyttäjien vastuun tunnistamiseen eettisten ja mahdollisimman vähän haittaa aiheuttavien tekoälyjärjestelmien varmistamisessa sekä tekoälyjärjestelmien vastuullisuutta edistävien mekanismien käyttöönoton ja toteuttamisen reaali maailman vaikutusten tunnistamiseen.

Tämän kokonaisuuden osaamistavoitteet:

- **Vastuullisuuden ymmärtäminen ja vastuullisuuden vaaliminen tekoälyssä:** Tässä kokonaisuudessa keskitytään tekoälyn vastuullisuuden peruskäsitteeseen, johon liittyy odotukset siitä, että suunnittelijat, kehittäjät ja käyttöönottajat noudattavat standardeja ja lainsäädäntöä, joilla varmistetaan tekoälyn asianmukainen toiminta koko sen elinkaaren ajan (Fjeld et al., 2020). Esitellään tekoälyn kehittäjien ja käyttäjien vastuuta eettisten ja mahdollisimman vähän haittaa aiheuttavien tekoälyjärjestelmien edistämisessä ja korostetaan, että on tärkeää kasvattaa vastuuntuntoa tekoälyn eettisestä kehittämisestä ja käytöstä ja arvostaa vastuullisuuden merkitystä haittojen minimoimisessa ja eettisen käyttäytymisen edistämisessä tekoälyjärjestelmissä.

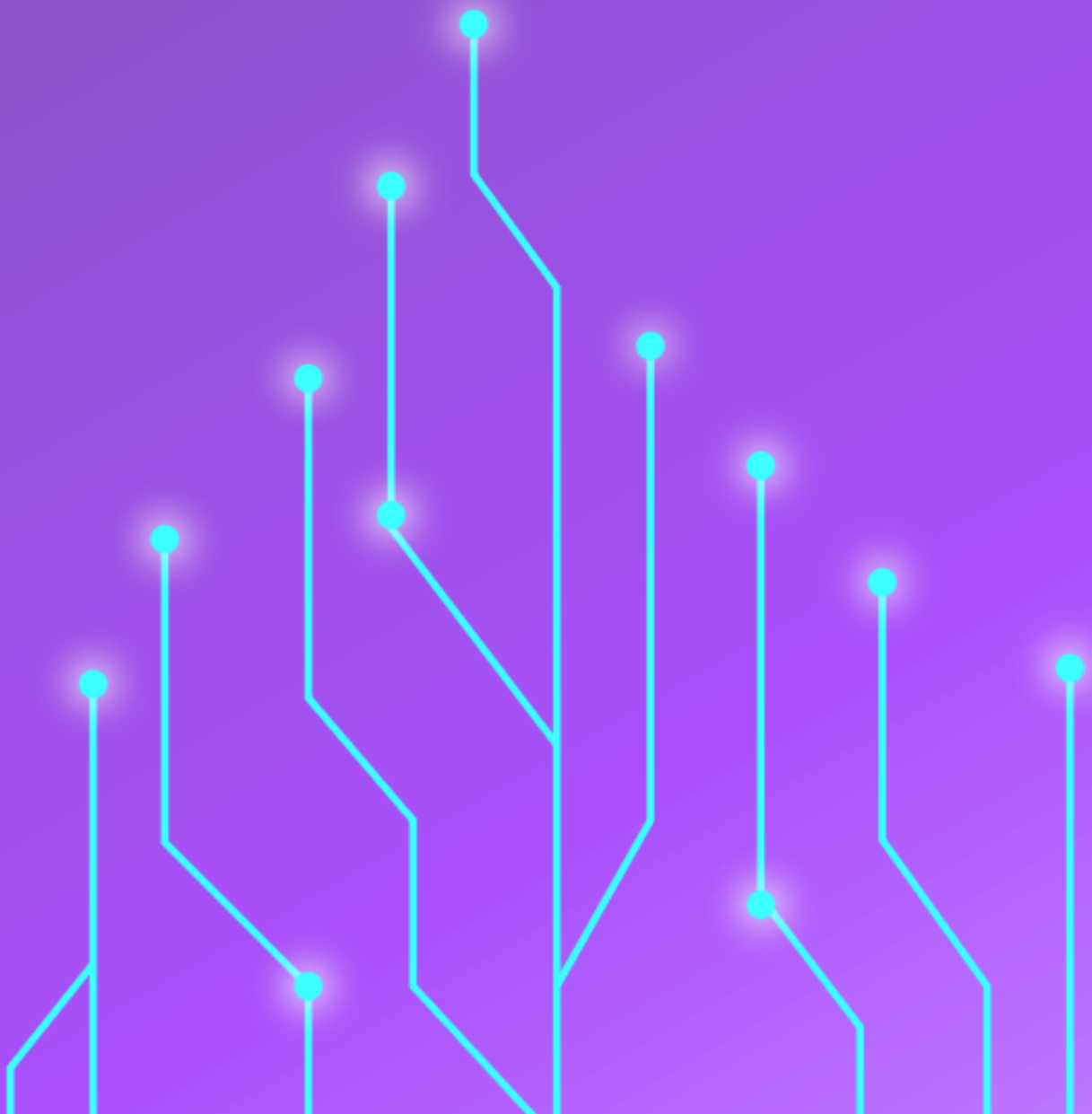


- **Vastuullisuuden rooli algoritmisen vinouman käsittelyssä**, jossa tunnustetaan vastuullisuuden keskeinen rooli algoritmisen vinouman vaikutusten ehkäisemisessä ja lieventämisessä. Tutkimme vastuullisuuden ja algoritmisen vinouman välistä suhdetta ja arvostamme sitä, että on tärkeää ottaa vastuu tekoälyjärjestelmistä päätöksenteon tehostamiseksi ja oikeudenmukaisempien tulosten edistämiseksi.
- **Vastuumekanismien toteuttaminen tekoälyjärjestelmissä**, jotta perehdytään erilaisiin mekanismeihin, joilla pyritään varmistamaan vastuullisuus tekoälyjärjestelmissä. Tutustumme tehokkaan viestinnän merkitykseen sidosryhmien kanssa, esimerkiksi antamalla tietoa tekoälyjärjestelmien perusteluista ja logiikasta, kuten syötteestä, tuotoksista ja päätöksentekoprosesseista. Esimerkkejä tällaisista mekanismeista ovat ohjeet, verkkoetiikka, hyväksyttävän käytön periaatteet (AUP) ja tekoälyn kehittäjille ja käyttäjille annetut määräykset.



02. Vastuullisuuden ymmärtäminen ja vastuullisuuden vaaliminen tekoälyssä

Osaamiskokonaisuus 3 | Vastuullisuus





02. Vastuullisuuden ymmärtäminen ja vastuullisuuden vaaliminen tekoälyssä

Vastuullisuus on olennaisen tärkeää tekoälyjärjestelmien eettisessä kehittämisessä ja käyttöönotossa.

Se toimii turvana epäeettisiä käytäntöjä vastaan ja varmistaa, että tekoälyteknologiat toimivat vakiintuneiden standardien, säännösten ja eettisten puitteiden mukaisesti. Tässä jaksossa perehdytään tekoälyn vastuullisuuden käsitteeseen ja tarkastellaan sen periaatteita, mekanismeja ja vaikutuksia nykypäivän teknologiamaisemissa.

Tekoälyn vastuullisuuteen liittyy pohjimmiltaan suunnittelijoiden, kehittäjien ja käyttöönottajien vastuu eettisten normien noudattamisesta tekoälyjärjestelmien koko elinkaaren ajan. Siihen kuuluu päätöksenteon avoimuus, algoritmien tuotosten jäljitettävyys ja sidosryhmien palautteen huomioon ottaminen. Lisäksi vastuullisuus edellyttää ennakoivaa riskien tunnistamista ja vähentämistä, jotta voidaan edistää tekoälyn eettistä kehittämistä ja käyttöä.

Oikeudelliset puitteet, alan standardit ja eettiset ohjeet muokkaavat tekoälyn vastuullisuutta. Oikeudellinen vastuuvastuu risteää sääntelyvaatimusten ja vastuukysymysten kanssa, mikä asettaa velvoitteita noudattaa tietosuojasäädöksiä ja kuluttajansuojasäädöksiä.



Alan standardit ja eettiset ohjeet tarjoavat puitteet eettisten ongelmien ratkaisemiseksi ja vastuullisten tekoälykäytäntöjen edistämiseksi.

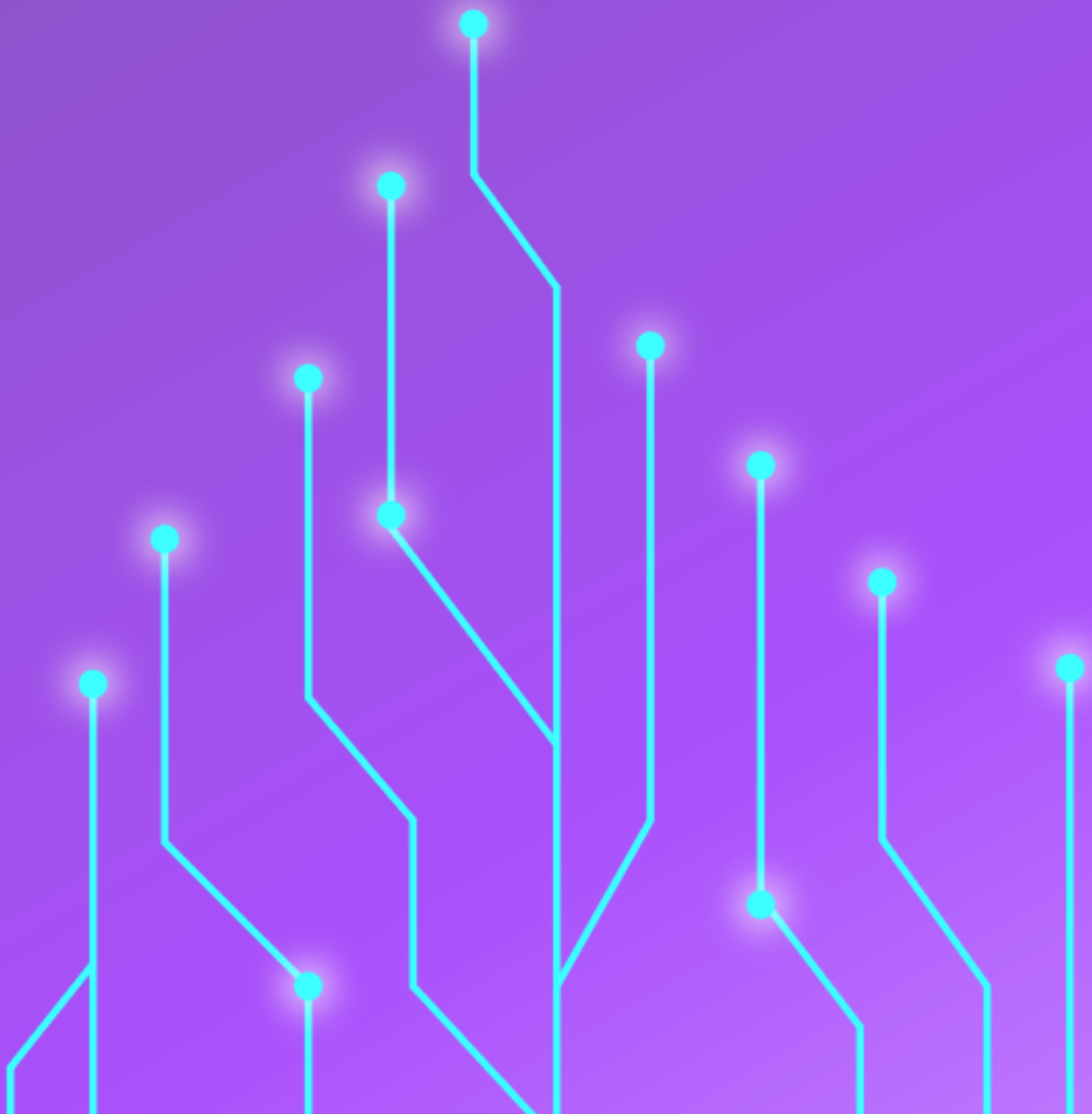
Algoritmiseen vastuullisuuteen kuuluu tekoälyalgoritmien teknisen tarkkuuden arviointi sekä niiden sosioekonomisten vaikutusten ja eettisten vaikutusten ymmärtäminen. Tekoälyalgoritmeissa esiintyvä vinouma voi johtaa epätasa-arvoisiin tuloksiin, jotka ylläpitävät eriarvoisuutta ja heikentävät luottamusta. Näin ollen vastuullisuus edellyttää valppautta vinoumien havaitsemisessa ja lieventämisessä sekä avoimuuden ja selitettävyyden mekanismeja.

Läpinäkyvyys ja selitettävyys lisäävät luottamusta ja vastuullisuutta tekoälyjärjestelmiin, sillä ne antavat sidosryhmille mahdollisuuden ymmärtää algoritmisen päätöksenteon prosesseja ja tuloksia. Eettinen vastuullisuus edellyttää tekoälyn kehittämisen yhdenmukaistamista eettisten periaatteiden, ihmisoikeuksien ja yhteiskunnallisten arvojen kanssa. Yhteiskunnallinen vastuullisuus edellyttää sidosryhmien sitouttamista sen varmistamiseksi, että tekoälyteknologiat kunnioittavat ihmisarvoa, itsemääräämisoikeutta ja oikeudenmukaisuutta.



03. Vastuullisuus: Algoritmiseen vinoumaan puuttuminen

Osaamiskokonaisuus 3 | Vastuullisuus





03. Vastuullisuus: Algoritmiseen vinoumaan puuttuminen

Vastuullisuudella on keskeinen rooli tekoälyjärjestelmien algoritmisiin vinoumiin puuttumisessa.

Kun tekoälyteknologia integroituu yhä enemmän yhteiskunnan eri osa-alueisiin, huoli vinoumasta ja oikeudenmukaisuudesta on noussut esiin. Algoritmisella vinoumalla tarkoitetaan tekoälyalgoritmeissa esiintyviä järjestelmällisiä virheitä tai epäoikeudenmukaisuutta, jotka johtavat tiettyjä yksilöitä tai ryhmiä syrjiviin tuloksiin. Nämä vinoumat voivat ylläpitää olemassa olevaa eriarvoisuutta, vahvistaa stereotypioita ja heikentää luottamusta tekoälyjärjestelmiin. Tässä osiossa tarkastelemme, miten vastuullisuus voi auttaa tunnistamaan, lieventämään ja ehkäisemään algoritmisiä vinoumia ja varmistamaan, että tekoälyteknologiat ovat oikeudenmukaisia, eettisiä ja tasapuolisia.

Yksi tärkeimmistä vastuullisuuteen liittyvistä näkökohdista algoritmiseen vinoumaan puuttumisessa on **läpinäkyvyys**. Läpinäkyvyydellä tarkoitetaan tekoälyjärjestelmien avoimuutta ja saavutettavuutta, jotta sidosryhmät voivat ymmärtää, miten algoritmit toimivat, miksi tietyt päätökset tehdään ja mitkä tekijät vaikuttavat niiden tuloksiin. Läpinäkyvät tekoälyjärjestelmät antavat käyttäjille mahdollisuuden tarkastella ja kyseenalaistaa algoritmien tuloksia, mikä edistää luottamusta ja vastuullisuutta.



Jos tekoälyjärjestelmää käytetään esimerkiksi automaattiseen päätöksentekoon lainojen myöntämisessä, avoimuus tarkoittaa, että päätöksentekoprosessissa huomioon otettavat kriteerit ja tekijät, kuten luottotiedot, tulot ja demografiset tiedot, julkistetaan. Avoimuuden ansiosta yksityishenkilöt voivat ymmärtää, miksi heidän lainahakemuksensa hyväksyttiin tai hylättiin, ja he voivat turvautua oikeussuojakeinoihin, jos he katsovat päätöksen olleen vinoutunut tai epäoikeudenmukainen.

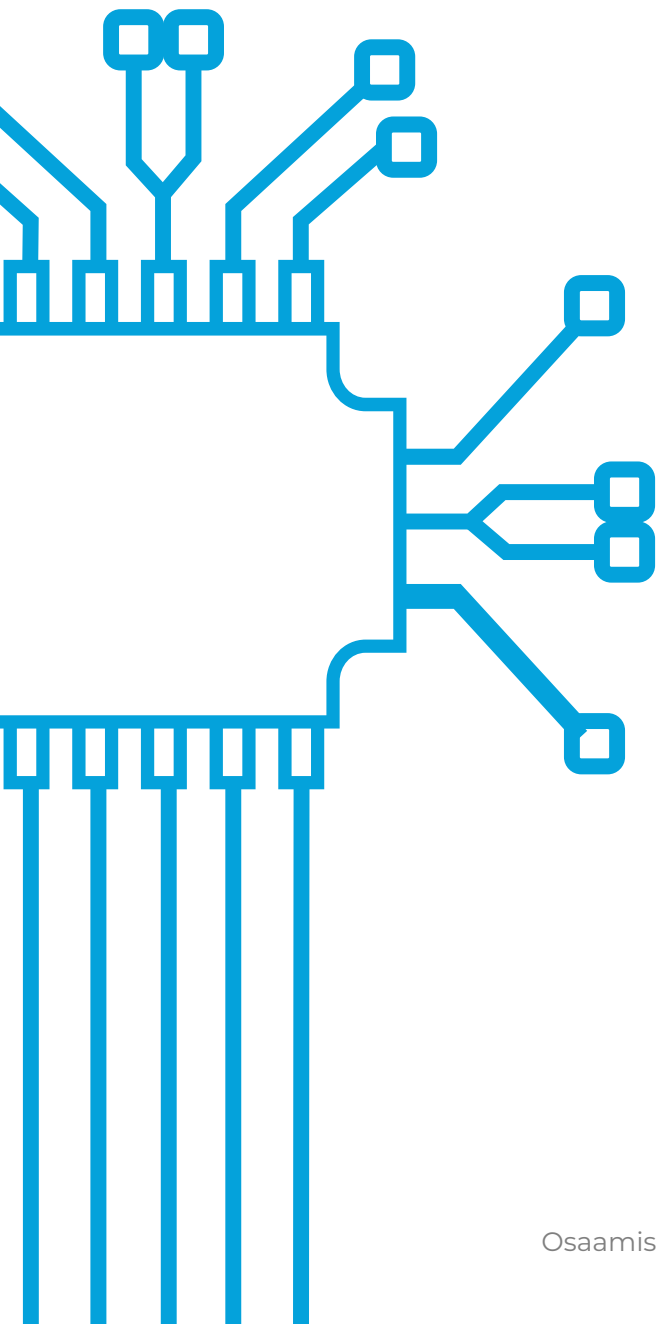
Toinen vastuullisuuteen liittyvä näkökohta algoritmiseen vinoumaan puuttumisessa on **selitettävyyys**. Selitettävyydellä tarkoitetaan tekoälyjärjestelmien kykyä antaa tulkinnanvaraisia selityksiä päätöksilleen ja toimilleen. Selitettävissä olevan tekoälyn avulla käyttäjät voivat ymmärtää algoritmien tulosten taustalla olevat perustelut ja tunnistaa ja korjata vinoumat tai virheet. Jos esimerkiksi tekoälyllä toimiva rekrytointityökalu hylkää automaattisesti sellaisten henkilöiden työhakemukset, joilla on tiettyjä demografisia ominaisuuksia, selitettävyyys tarkoittaa selitysten antamista siitä, miksi kyseiset hakemukset hylättiin, esimerkiksi korostamalla erityisiä kriteerejä tai ominaisuuksia, jotka johtivat päätökseen. Selitettävyyys antaa käyttäjille mahdollisuuden kyseenalaistaa vinoutuneet päätökset ja ryhtyä korjaaviin toimiin tekoälyjärjestelmien vinouman lieventämiseksi.

Vastuullisuuteen kuuluu myös **vastuu tekoälyjärjestelmien suunnittelusta, kehittämisestä ja käyttöönnotosta**. Tekoälyn kehittäjillä ja ammattilaisilla on vastuu varmistaa, että heidän järjestelmänsä ovat oikeudenmukaisia, puolueettomia ja tasapuolisia. Tähän vastuuseen kuuluu perusteellinen testaus ja validointi tekoälyalgoritmien vinoumien tunnistamiseksi ja lieventämiseksi sekä tekoälyjärjestelmien seuranta ja auditointi sen varmistamiseksi, että ne noudattavat jatkuvasti eettisiä ja oikeudellisia normeja. Tekoälyn kehittäjät voivat esimerkiksi käyttää tekniikoita, kuten vinoumien havaitsemisalgoritmeja, oikeudenmukaisuusmittareita ja vastakohtaista testausta, malleissaan esiintyvien vinoumien tunnistamiseksi ja poistamiseksi. Ne voivat myös ottaa käyttöön prosesseja tekoälyjärjestelmien säännöllisiä tarkastuksia ja tarkistuksia varten, jotta ajan mittaan mahdollisesti esiin nousevat vinoumat voidaan tunnistaa ja käsitellä.

Lisäksi vastuullisuus algoritmiseen vinoumaan puuttumisessa edellyttää **yhteistyötä ja sitoutumista eri sidosryhmien kanssa**. Tähän kuuluu, että tekoälyjärjestelmien suunnitteluun, kehittämiseen ja valvontaan otetaan mukaan yhteisöt, joita asia koskee, etujärjestöt, poliittiset päättäjät ja sääntelyelimet. Sidosryhmien kanssa yhteistyötä tekemällä, tekoälyn ammattilaiset voivat saada arvokasta tietoa teknologioidensa mahdollisista vaikutuksista ja varmistaa, että heidän järjestelmänsä suunnitellaan ja otetaan käyttöön siten, että kaikkien yksilöiden ja ryhmien oikeuksia ja etuja kunnioitetaan.

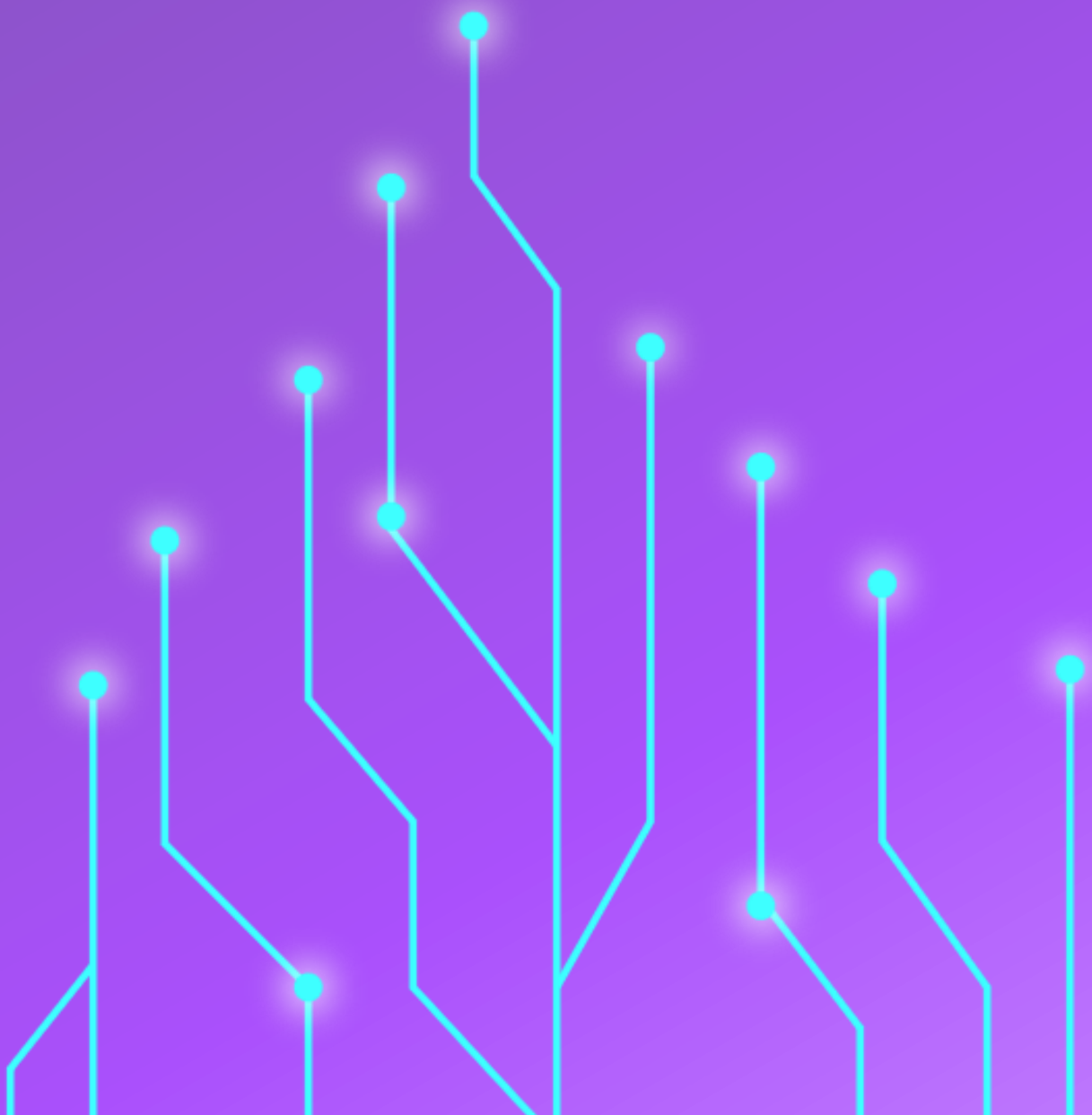


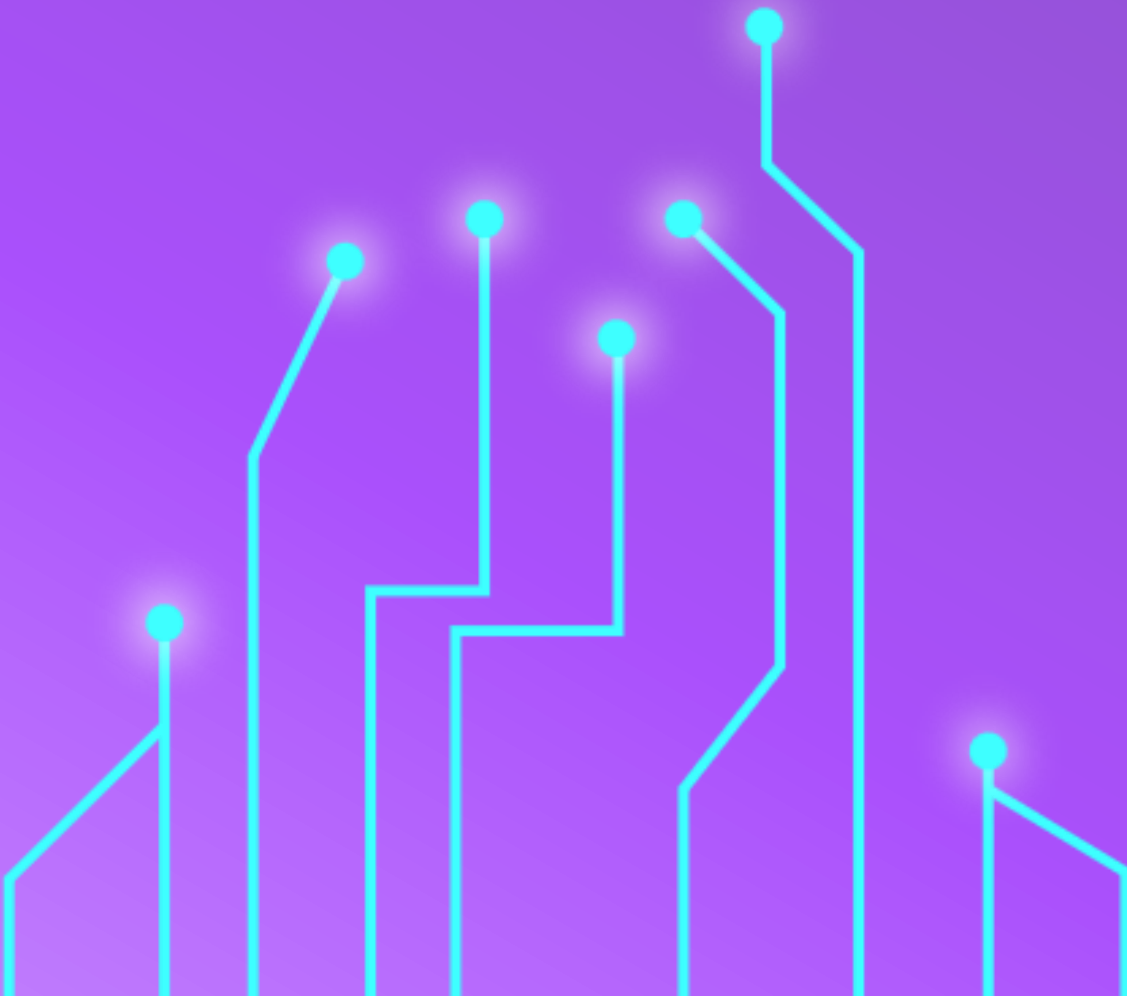
Jos tekoälyjärjestelmää käytetään esimerkiksi ennakoivaan poliisitoimintaan, vastuullisuuteen kuuluu yhteisön jäsenten ja kansalaisoikeusjärjestöjen kuuleminen, jotta voidaan ymmärtää niiden huolenaiheet ja näkemykset tekoälyn käytöstä lainvalvonnassa ja puuttua mahdollisiin vinoumiin tai haittoihin.



04. Vastuumekanismit, joilla puututaan algoritmiseen vinoumaan

Osaamiskokonaisuus 3 | Vastuullisuus





04. Mekanismit vastuuvollisuuden varmistamiseksi tekoälyjärjestelmissä

Nykyisessä nopeasti kehittyvässä teknologisessa ympäristössä, jossa tekoälyjärjestelmät vaikuttavat merkittävästi päätöksentekoprosesseihin, vastuullisuuden varmistaminen on ensiarvoisen tärkeää. Mekanismit, joilla varmistetaan tekoälyjärjestelmien vastuullisuus, käsittävät erilaisia välineitä ja kehyksiä, joiden tarkoituksena on edistää eettistä käyttäytymistä, avoimuutta ja oikeudenmukaisuutta tekoälyteknologian kehittämisessä ja käytössä:

- **Ohjeet** toimivat tekoälyn kehittäjien ja käyttäjien yleisinä periaatteina ja parhaina käytänteinä, jotka tarjoavat ohjeita eettisistä näkökohdista, vastuullisesta tietojenkäsittelystä ja riskinhallintastrategioista.
- **Netiketti**- eli internet-etikettiperiaatteita voidaan mukauttaa tekoälyjärjestelmien eettisen käyttäytymisen edistämiseen korostamalla kunnioitusta, kohteliaisuutta ja vastuullista viestintää käyttäjien välillä.
- **Hyväksyttävää käyttöä koskevissa toimintalinjoissa (Acceptable Use Policies, AUP)** hahmotellaan tekoälyjärjestelmiä koskevat käyttäytymisnormit ja käyttöohjeet, joissa määritellään hyväksyttävä käyttäytyminen ja väärinkäytön seuraukset.



- **Säätelykehyksissä** vahvistetaan tekoälyn kehittämistä ja käyttöönottoa koskevat oikeudelliset vaatimukset ja standardit, mukaan lukien vastuuvollisuus, avoimuutta ja valvontaa koskevat säännökset. Euroopan unionin (EU) **säätelykehyksissä, kuten yleisessä tietosuoja-asetuksessa (GDPR)** ja ehdotetussa **tekoälylaissa**, vahvistetaan tekoälyn kehittämistä ja käyttöönottoa koskevat oikeudelliset vaatimukset ja standardit. Näihin säädöksiin sisältyy vastuullisuutta koskevia säännöksiä, kuten vaatimuksia tekoälyjärjestelmien kehittämisen ja käytön läpinäkyvyydestä, oikeudenmukaisuudesta ja valvonnasta. Esimerkiksi yleisen tietosuoja-asetuksen mukaan kehittäjien on varmistettava avoimuus siinä, miten tekoälyjärjestelmät käsittelevät henkilötietoja, ja annettava yksilöille tietoa automaattiseen päätöksentekoon liittyvästä logiikasta.

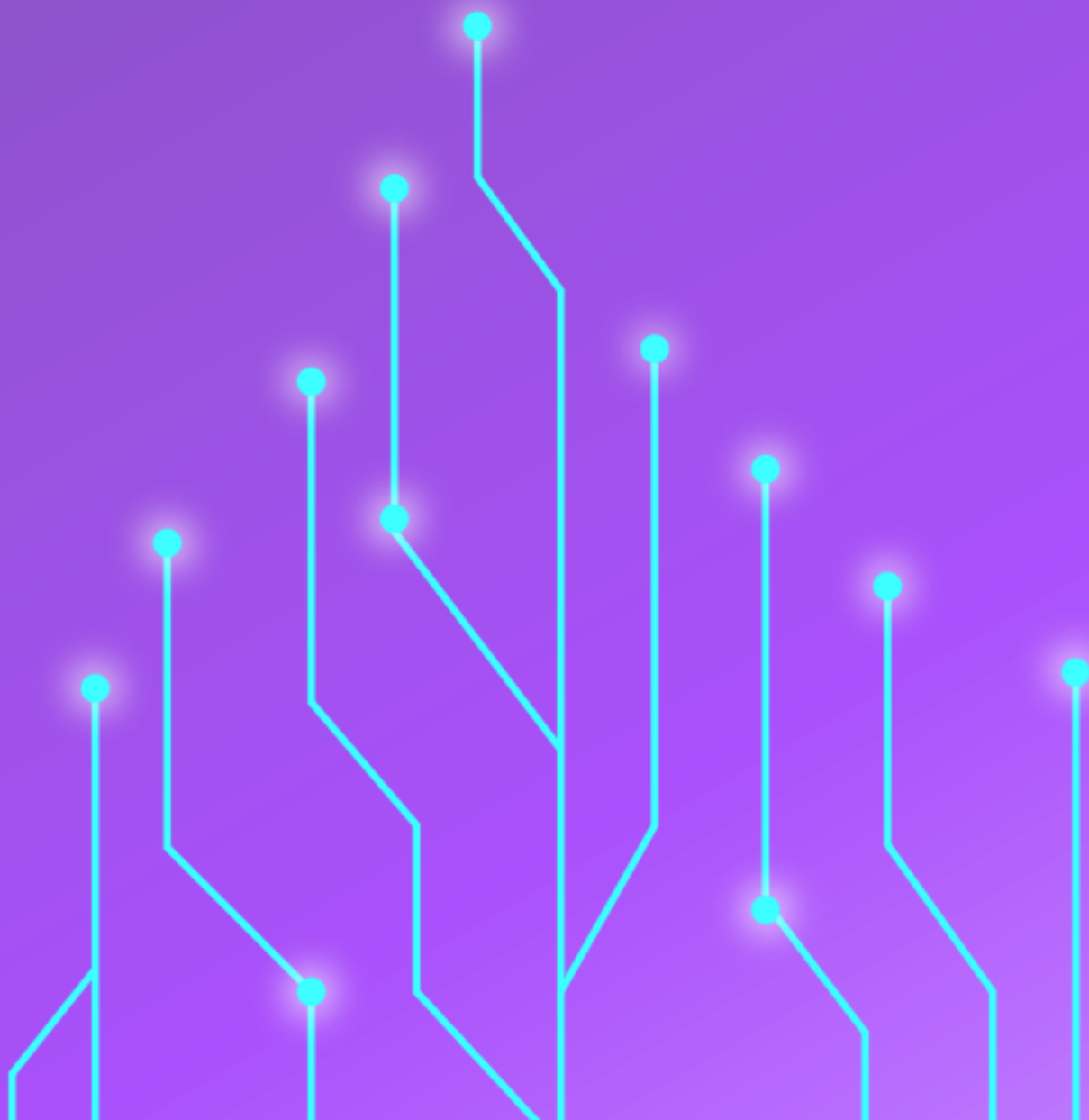


Näitä mekanismeja hyödyntämällä sidosryhmät voivat edistää vastuullisuuden ja eettisen käyttäytymisen kulttuuria tekoälyn kehittämisessä ja käyttöönotossa ja varmistaa, että tekoälyteknologiaa käytetään vastuullisesti ja eettisesti yhteiskunnan hyväksi.



05. Yhteenveto

Osaamiskokonaisuus 3 | Vastuullisuus





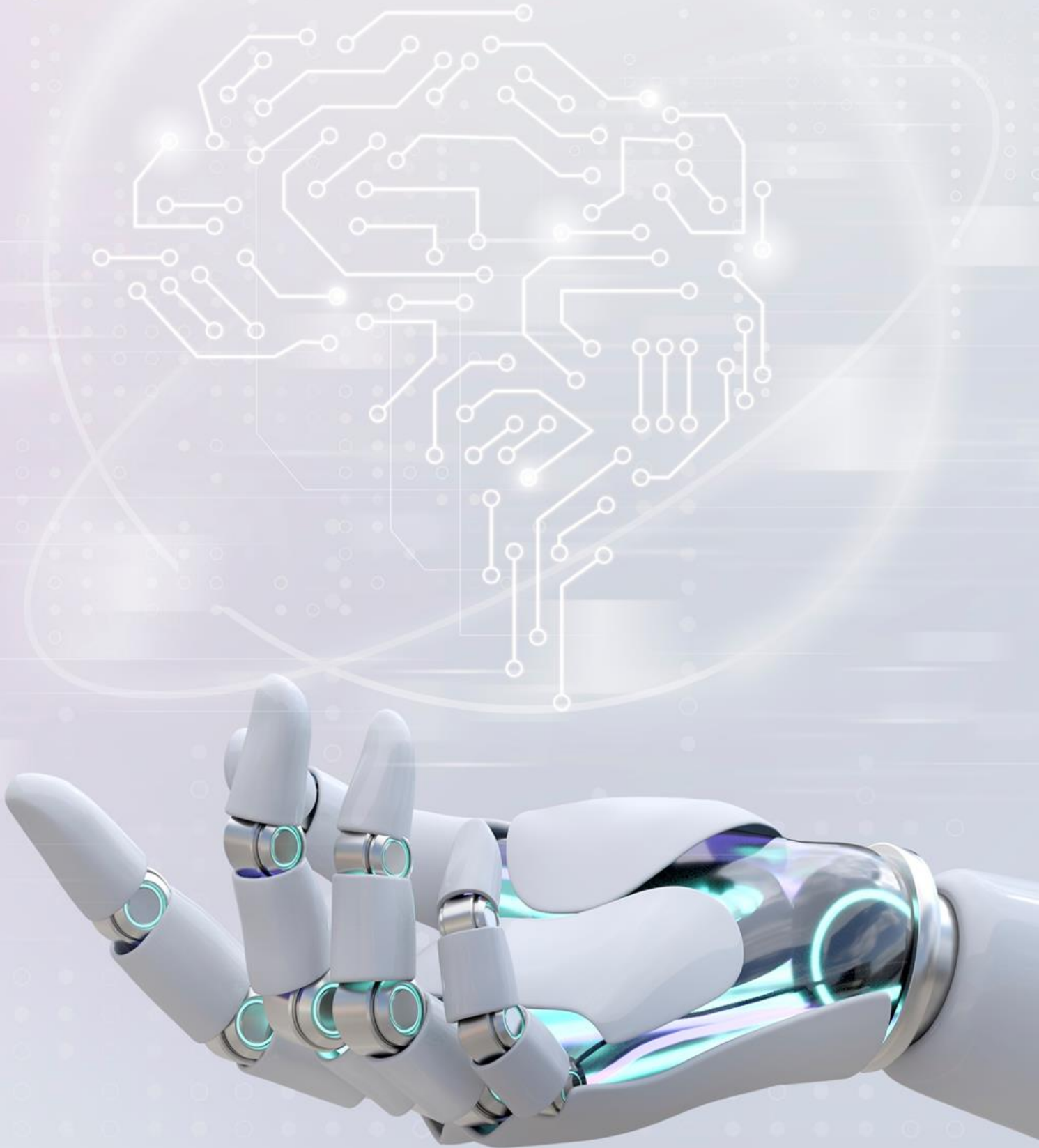
05. Yhteenveto

Yhteenvetona voidaan todeta, että vastuullisuus on tekoälyjärjestelmien eettisen kehittämisen ja käyttöönoton kulmakivi. Olemme tutkineet tässä osaamiskokonaisuudessa tekoälyn vastuullisuuden monitahoista luonnetta ja perehtyneet sen peruseriaatteisiin, taustalla oleviin mekanismeihin ja käytännön vaikutuksiin. Läpinäkyvyydestä ja selitettävyydestä sidosryhmien sitoutumiseen ja sääntelyn noudattamiseen vastuullisuus on johtava periaate, jolla varmistetaan, että tekoälyteknologioita kehitetään ja käytetään tavalla, joka noudattaa eettisiä normeja, edistää avoimuutta ja edistää oikeudenmukaisuutta ja tasapuolisuutta.

Vastuullisuuden avulla sidosryhmät voivat selviytyä tekoälyteknologian monimutkaisista kysymyksistä rehellisesti, vastuullisesti ja sitoutumalla yhteiskunnalliseen hyvinvointiin.

Kun tekoäly kehittyy edelleen ja läpäisee ihmiselämän eri osa-alueet, vastuullisuus on edelleen ensiarvoisen tärkeää sellaisen tulevaisuuden muotoilussa, jossa tekoälyteknologiat palvelevat yhteistä hyvää ja vaalivat eettisiä, oikeudenmukaisia ja ihmisarvoisia arvoja.







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**



ISQe
ENGAGING PEOPLE



INNOVATION TRAINING CENTER



AARHUS UNIVERSITY



VAMK
AMMATTIOPETTAJAKORKEAKOULU
UNIVERSITY OF APPLIED SCIENCES

helixconnect



2022-1-ES01-KA220-HED-000085257