



CHARLIE – Massadatan vinouman haaste tekoälyn ja koneoppimisen opetuksessa

Projektinnumero: 2022-1-ES01-KA220-HED-
000085257

Työpaketti 2 Kompetenssimatriisit

Tämä hanke on rahoitettu Euroopan komission tuella. Tämä suunnitelma kuvastaa ainoastaan sen laatijan näkemyksiä, eikä komissiota voida pitää vastuullisena sen sisältämien tietojen mahdollisesta käytöstä.

Asiakirjan otsikko	CHARLIE-hanke: WP2 "Algoritminen ennakkoluuloisuus" - osaamismatriisit.
Myöntämispäivä	29/09/2023 (versio 4.0)
Kirjoittaja(t)	Ruben Comas (UIB)
Sähköpostiosoite	rubencomas@uib.es
Asiakirjaan osallistuneet tahot	Kaikki kumppanit
Laatutarkastaja (jos sellainen on)	Sidosryhmät
Sivujen lukumäärä	76
Luottamuksellisuuden tila	CHARLIE Vain yhteistyökumppanit ja kutsutut asiantuntijat

Sisällysluettelo

1. JOHDANTO CHARLIE-PROJEKTIIN	5
1.1. Charlie-projektin yleiskatsaus	5
1.2 Projektin tavoitteet	6
1.3. Charlie-projektin kunkin työpaketin odotetut tulokset kootusti.....	6
1.4. CHARLIEn kohderyhmät	7
1.5. Työpaketti 2:n tavoitteet.....	8
1.6. Työpaketti 2:n odotetut tulokset ja osallistujat.....	9
2. JOHDANTO: ALGORITMIT, TEKOÄLY, VINOUMAT JA ETIIKKA.....	10
3. KOMPETENSSIMATRIISI: TEOREETTINEN LÄHESTYMISTAPA.....	12
4. CHARLIE TYÖPAKETTI 2 -MATRIISIT.....	16
4.1. ALGORITMIN VINOUMA -KURSSI EQF6.....	17
4.1.1. Algoritmin vinouma -kurssin sisältö.....	19
4.1.2. Kohderyhmän kuvaus, EQF6-ominaisuudet ja piirteet.....	20
4.1.3. Algoritmin vinouma -kurssin yleinen kompetenssimatriisi EQF6.....	22
4.1.4. Kompetenssimatriisi kullekin Algoritmin vinouma -kurssin osaamisyksikölle EQF6.....	24
4.2. CHARLIE Työpaketti 2 - Eettisen tekoälyn mikrokredentiaalikurssi EQF4	51
4.2.1. Eettisen tekoälyn mikrokredentiaalikurssin EQF4 sisältö	52
4.2.2. Kohderyhmän kuvaus, EQF4-ominaisuudet ja piirteet.....	53
4.2.3. Yleinen kompetenssimatriisi EQF4:n mikrokredentiaalikurssia varten	56
4.2.4. EQF4-mikrokredentiaalikurssin kunkin osaamisyksikön osaamismatriisi.....	57

4.3. CHARLIE WP2:n oppimistulokset aikuisille ja nuorille EQF2 (SERIOUS GAME).....	70
4.3.1. Algoritmisen vinouman pohdintapelin sisältö.....	71
4.3.2. Kohderyhmän kuvaus, EQF2-ominaisuudet ja piirteet	72
4.3.3. Yleinen kompetenssimatriisi Algoritmisen vinouman pohdintapelille EQF2:ssa.....	73
Viitteet.....	75

1. JOHDANTO CHARLIE-PROJEKTIIN

1.1. Charlie-projektin yleiskatsaus

CHARLIE on ERASMUS+ KA2 -hanke, jonka toteutusaika on 30 kuukautta, 30.12.2022-29.6.2025. Hankkeen toteuttaa konsortio, johon kuuluu kuusi (6) kumppania viidestä (5) Euroopan maasta: Espanja, Portugali, Romania, Suomi ja Tanska.

Tekoäly (AI) on jokapäiväistä elämäämme. Tekoäly on kaikkialla. Niin algoritmissa, joka tunnistaa kasvomme kadulla kävellessämme ja syöttää poliisin biometrisiä turvapalveluja, kuin myös algoritmissa, joka valitsee sosiaalisessa mediassa näkemämme mainokset. Mutta vaikka koneoppiminen ja tekoäly ovat matematiikkaa, eivät ne aina ole oikeassa, sillä tiedot, joita päätelmien tekemiseksi käsitellään, voivat olla, ja usein ovatkin, puolueellisia. Yhteiskuntatieteissä on tutkittu ihmisen puolueellisuutta jo vuosia. Se syntyy implisiittisistä assosiaatioista, jotka heijastavat ennakkoluuloja, joista emme ole tietoisia ja jotka voivat johtaa moniin kielteisiin lopputuloksiin. Tekoälyä ja koneoppimista ei ole suunniteltu tekemään eettisiä päätöksiä. Se ei ole eettinen algoritmi. Se tekee ennusteita, jotka perustuvat siihen, miten maailma toimii nykyään, ja edistää siten osaltaan ennakkoluuloja ja syrjiviä käytäntöjä, jotka ovat systemaattisesti juurtuneet tämän päivän yhteiskuntiimme. Kun tekoälyn ja koneoppimisen teknologioita käytetään laajalti, ja ne ovat usein suurten teknologiayhtiöiden omistuksessa, joiden ainoa tavoite on tuottaa voittoa, on teknologiaan saatava pikaisesti ihmiskeskeinen lähestymistapa, ja sitä on käytettävä sosiaalisten ongelmien ratkaisemiseen sen sijaan, että se edistäisi niitä. Euroopan komissio esitti 25.4.2018 ja 7.12.2018 antamissaan tiedonannoissa tekoälyä koskevan visionsa, jossa tuetaan "eettistä, turvallista ja huippuluokan tekoälyä, joka tehdään Euroopassa". Tekoälyjärjestelmien on oltava ihmiskeskeisiä, ja niiden on perustuttava sitoumukseen käyttää niitä ihmiskunnan ja yhteisen hyvän palveluksessa, tavoitteena parantaa ihmisten hyvinvointia ja vapautta. Vaikka tekoälyjärjestelmät tarjoavat suuria mahdollisuuksia, liittyy niihin myös tiettyjä riskejä, joita on käsiteltävä asianmukaisesti ja oikeasuhtaisesti. Meillä on nyt merkittävä tilaisuus muokata niiden kehitystä.

CHARLIE pyrkii varmistamaan, että voimme luottaa niihin sositelknisiin ympäristöihin, joihin ne on upotettu. Lisäksi toiveenamme on, että tekoälyjärjestelmien tuottajat saavat kilpailuetua sisällyttäessään luotettavan tekoälyn tuotteisiinsa ja palveluihinsa. Tämä tarkoittaa sitä, että pyritään maksimoimaan tekoälyjärjestelmien hyödyt ja samalla ehkäisemään ja minimoimaan niiden riskejä. Korkea-asteen koulutus, aikuiskoulutus ja nuorten koulutus tarvitsevat uusia ja innovatiivisia opetussuunnitelmia, joilla

voidaan vastata tähän osaamisvajeeseen ja antaa oppijoille tiedot ja taidot, joiden avulla he voivat edistää eettisempää lähestymistapaa teknologian kehittämiseen. Tarve tehdä teknologiakasvatuksesta inhimillisempää on linjassa digitaalisen koulutuksen toimintasuunnitelman kanssa, joka sisältää erityistoimia, joilla käsitellään tekoälyn ja datan käytön eettisiä vaikutuksia ja haasteita koulutuksessa.

1.2 Projektin tavoitteet

CHARLIEn tavoitteena on vastata massadatan vinouman haasteeseen edistämällä eettisiä ja osallistavia lähestymistapoja teknologian, tekoälyn ja koneoppimisen opetuksessa. CHARLIEn päätavoitteet ovat seuraavat:

1. Lisätä korkeakoulujen mahdollisuuksia tarjota opiskelijoille verkko-oppimismahdollisuuksia, jotka vastaavat yhteiskunnan tarpeisiin, ja jotka ovat myös räätälöity opiskelijoiden oppimistarpeisiin;
2. Kehittää tekniikan opiskelijoiden sosiaalista ja eettistä osaamista, voidakseen suhtautua tekoälyn ja koneoppimisen teknologioihin myönteisesti, kriittisesti ja eettisesti;
3. Lisätä opetushenkilökunnan valmiuksia hyödyntää tehokkaasti digitaalisia ja osallistavia menetelmiä tekoälyn ja koneoppimisen opetuksessa (erityisesti verkko opetuksessa);
4. Luoda synergioita korkeakoulujen, aikuiskoulutuksen ja nuorten (12–18-vuotiaat) opetuksen välille tekoälyn etiikan opetuksessa;
5. Mahdollistaa korkeakoulujen kurssien siirrettävyyttä tekoälyn vinoumassa aikuiskoulutukseen ja ammatilliseen koulutukseen;
6. Lisätä tietoisuutta tekoälyn datan vinoumasta yhteiskunnallisella tasolla.

1.3. Charlie-projektin kunkin työpaketin odotetut tulokset kootusti

1- Kompetenssimatriisit kursseille "Kurssi algoritmin vinoumasta" (EQF6), "Eettisen tekoälyn mikrokredentiaalikurssi" (EQF4) ja virtuaaliselle hyötypelille (EQF2) **(Työpaketti 2).**

2- Kurssi algoritmin vinoumasta (korkea-aste) – Oppimisen lähestymistavat: a) synkroniset istunnot (mieluiten kasvokkain), b) Verkko-oppimisen asynkroniset istunnot teoreettista osuutta varten, täydennettynä virtuaalisella hyötypelillä formatiivisena arviointina ja verkossa tapahtuvalla summatiivisella arvioinnilla, joka toimitetaan .scorm-muodossa valmiina ladattavaksi mihin tahansa oppimisen hallintajärjestelmään opiskelijoille, jotka opiskelevat massadatan, tekoälyn, koneoppimisen ja syväoppimisen kursseja. **(Työpaketti 3)**

- 3- Työkalupakki synkroniseen opetukseen** -
opettajille/professoreille/luennoitsijoille, jotka toteuttavat synkronisia istuntoja (kasvokkain tai verkossa) täydentämään sähköisen oppimisen asynkronisia istuntoja. **(Työpaketti 3)**
- 4 - Ohjeet digitaalisen koulutuksen tarjonnasta vastaavien yliopistojen hallinto- ja koulutusyksiköiden johdon valmiuksien lisäämiseksi**, jotta voidaan edistää digitaalisen koulutuksen tarjontaa algoritmien vinoumista. **(Työpaketti 3)**
- 5 - Webinaarit** edistämään vertaisoppimista ja pohtimaan korkeakoulujen roolia teknologiakasvatuksessa, edistämään yhteiskuntatieteiden monitieteisyyttä teknologiakasvatuksessa ja tarkastelemaan lähestymistapoja yhteentoimivuuteen nuorten koulutuksen ja aikuiskoulutuksen kanssa. **(Työpaketti 3)**
- 6 - Aikuisopiskelijoille suunnattu Eettisen tekoälyn mikrokredentiaalikurssi (EQF4)**, joka suoritetaan verkossa ja asynkronisesti. **(Työpaketti 4)**
- 7- Virtuaalinen hyötynä (EQF2)** 12-18-vuotiaille nuorille (pääasiassa heikommissa asemassa oleville/nuorille naisille) tarkoituksena edistää tasaista sukupuolijakaumaa STEM-alalla jo nuoresta iästä lähtien. **(Työpaketti4)**
- 8- Työkalupakki aikuisten ja nuorten tukemiseksi** eettisen tekoälyn täydennyskoulutuksessa. **(Työpaketti 4)**
- 9- Toimintapoliittinen suositus** aikuisopiskelijoiden mikrokredentiaalikurssin tunnustamiseksi heidän pyrkiessään korkea-asteen opintoihin tekniikan aloilla. **(Työpaketti 4)**

1.4. CHARLIEn kohderyhmät

CHARLIE-hankkeen tärkeimmät kohderyhmät ovat:

- **Korkeakoulut**, jotka tarjoavat massadataa, tekoälyyn, koneoppimiseen ja syväoppimiseen liittyviä oppimismahdollisuuksia/kursseja, sekä niiden hallintovirkamiehet/koulutusosastojen johto.
- **Korkeakouluopiskelijat**, jotka opiskelevat massadatan, tekoälyn, koneoppimisen ja syväoppimisen kursseja.
- **Korkeakouluopettajat/professorit/luennoitsijat** joko yhteiskuntatieteiden tai tekniikan alan aloilta.
- **Aikuiskoulutus**
 - Kaikenikäiset ja eri sosioekonomisista lähtökohdista tulevat aikuiset, jotka pyrkivät saavuttamaan korkeamman pätevyytason, jolla on merkitystä työmarkkinoilla ja aktiivisessa yhteiskunnallisessa osallistumisessa.
- **Nuorisjärjestöt** ja niiden henkilöstö
- **Nuoret** (12–18-vuotiaat) - erityisesti nuoret naiset ja sosiaalisesti heikommissa asemassa olevat nuoret (koulutusjärjestelmään liittyvät esteet, kulttuurierot,

sosiaaliset esteet, taloudelliset esteet, syrjinnän aiheuttamat esteet ja maantieteelliset esteet).

Viestinnän ja projektin levitystoimien ja -tuotteiden tärkeimmät kohderyhmät:

- Kansalliset sääntelyviranomaiset ja päätöksentekijät;
- Laajempi yleisö
- Korkeakoulut (yliopistojen opettajat, opiskelijat (opiskelijat ja jatko-opiskelijat) ja korkeakoulujen hallintoelimet).
- Aikuiskoulutuksen tarjoajat
- Nuorisjärjestöt
- Ammatillisen koulutuksen tarjoajat
- Tekoälyfoorumit ja -alustat
- Kansalaisjärjestöt, jotka puuttuvat erilaisiin vinoumiin ja syrjintään.
- Opetusteknologian tarjoajat
- Joukkotiedotusvälineiden toimijat (TV, sähköiset sanomalehdet, radiot jne.).
- Tekoälyn ja koneoppimisen pk-yritykset ja yritykset.
- Korkea-asteen koulutuksen järjestöt, mukaan lukien niihin liittyvät yhteistyöorganisaatiot, jotka edistävät parhaita käytäntöjä EU:n tasolla ja levittävät niitä lumipallon lailla korkeakoulujensa jäsenille (esim: European University Association, Science Business Network, Triple Helix Association, AMBA, University-Industry Innovation Network, European Business Angels Network (eettisiin tekoälyinnovaatioihin sijoittamiseen)).
- EU:n laajuiset opiskelijaelimet
- Osaamisen seuranta-/kehittämisyhdistykset (esim. DigCompEdu-yhteisöt).

1.5. Työpaketti 2:n tavoitteet

Kompetenssimatriisit "Algoritmin vinouma" -työpaketin erityistavoitteet ovat seuraavat:

- määritellä yhteinen näkemys korkeakouluopiskelijoille kehitettävän Algoritmin vinouma -kurssin oppimispolun tarkoituksesta ja oppimistavoitteista;
- laatia yhteinen näkemys oppimispolun tarkoituksesta ja oppimistavoitteista, joka kehitetään aikuisopiskelijoille Eettisen tekoälyn mikrokredentiaalikurssilla;
- laatia yhteinen näkemys eettistä tekoälyä käsittelevän nuorille kehitettävän hyötypelin oppimispolun tarkoituksesta ja oppimistavoitteista.

Nämä tavoitteet edistävät yleistavoitteita asettamalla kulmakivet seuraaville toimille, jotta opettajilla, ohjaajilla ja oppijoilla olisi selkeä näkemys eri oppimispolkujen tavoitteista, ja jotta korkeakoulujen, aikuiskoulutuksen ja

nuorten koulutuksen organisaatioiden valmiuksia tarjota yhteiskunnan tarpeita vastaavia oppimismahdollisuuksia, jotka on räätälöity oppijoiden oppimistarpeiden mukaan.

1.6. Työpaketti 2:n odotetut tulokset ja osallistujat

1.6.1. Kompetenssimatriisi EQF 6 Algoritmin vinouma -kurssille (korkeakouluopiskelijat) - Oppimistulosten lähestymistapa, 2 ECVS-pistettä, akkreditointisuositukset, ilmoittautumisen ennakko vaatimukset, kontaktitunnit, kokonaistyömäärä, EntreComp-pätevyyden (esim.: Eettinen ja kestävä ajattelu), DigComp 2.0 -pätevyyden (esim.: Terveiden ja hyvinvoinnin suojeleminen) ja GrenComp-pätevyyden (esim.: Oikeudenmukaisuuden tukeminen) integrointi.

Kompetenssiyksiköt:

CU1 - Algoritmien mallit ja rajoitukset UIB

CU2 - Tiedon oikeellisuus ja vääristymät tekoälyssä UA

CU3 - Tekoälyn yksityisyys ja soveltuvuus UA

CU4 - Tekoälyn etiikka, käytännön lähestymistapa VAMK

CU5 - Tapaustutkimukset ja hanke VAMK

1.6.2. Kompetenssimatriisi EQF 4 Eettisen tekoälyn mikrokredentiaalikurssia varten - Oppimistuloksia koskeva lähestymistapa, mahdolliset ECVS-pisteet, akkreditointisuositukset, ilmoittautumisen ennakko vaatimukset, kontaktitunnit, kokonaistyömäärä, EntreComp-pätevyyden (esim.: Eettinen ja kestävä ajattelu), DigComp 2.0 -pätevyyden (esim.: Terveiden ja hyvinvoinnin suojeleminen) ja GrenComp-pätevyyden (esim.: Oikeudenmukaisuuden tukeminen) integrointi.

Kompetenssiyksiköt:

CU1 - Mikä on algoritmin vinouma? HELIX

CU2 - Väärinkäyttö(non-maleficence) HELIX

CU3 - Vastuu ITC

CU4 - Läpinäkyvyys ITC

CU5 - Ihmisoikeudet ja oikeellisuus ITC

CU6 - Tekoälyn etiikka, käytännön lähestymistapa HELIX

1.6.3. Kompetenssimatriisi nuorille EQF 2 (hyötypeli) - Oppimistulosten lähestymistapa, ilmoittautumisen ennakko vaatimukset, kontaktitunnit, EntreComp-pätevyyden (esim.: Eettinen ja kestävä ajattelu), DigComp 2.0 -pätevyyden (esim.: Terveiden ja hyvinvoinnin suojeleminen) ja GrenComp-pätevyyden (esim.: Oikeudenmukaisuuden tukeminen) integrointi. **Yleinen sisältö:** Järjestelmällinen epätasa-arvo yhteiskunnassa, tekoälyjärjestelmien perusmekaniikka, poliittiset agendat, tekoälyn vaikutukset maailmassa.

2. JOHDANTO: ALGORITMIT, TEKÖÄLY, VINOUMAT JA ETIIKKA

Algoritmeilla on keskeinen rooli nyky-yhteiskunnassa, sillä ne auttavat erilaisissa palveluissa, kuten kappaleiden, elokuvien tai ystävien suositelussa (Paraschakis, 2018; Milano ym., 2020). Niitä käytetään myös kouluissa, sairaaloissa, rahoituslaitoksissa, tuomioistuimissa ja hallituksissa tärkeiden päätösten tekemiseen (Obermeyer ym., 2019). Algoritmien käyttö voi kuitenkin aiheuttaa eettisiä riskejä (Floridi ym., 2018).

Algoritmit voivat olla puolueellisia, mikä voi vaikuttaa esimerkiksi käännöksiin (Prates ym., 2020) tai työpaikkailmoituksiin (Lambrecht ja Tucker, 2019). Esimerkiksi paremmin palkattuja työpaikkoja saatetaan näyttää miehille enemmän kuin naisille. Tämä voi johtaa myös epäoikeudenmukaiseen kohteluun terveydenhuollossa, kuten siihen, että valkoihoiset potilaat saavat parempaa hoitoa kuin tummaihoiset potilaat (Obermeyer ym. 2019). Vaikka ihmiset yrittävät korjata näitä ongelmia, uusia ongelmia tulee jatkuvasti esiin.

Vuodesta 2012 lähtien tekoälyyn ja sen eettisiin vaikutuksiin on kiinnitetty paljon huomiota (Perrault ym. 2019). Vuonna 2016 eräessä tutkimuksessa pyrittiin kartoittamaan näitä huolenaiheita (Mittelstadt ym. 2016), mutta ala on kasvanut ja muuttunut paljon sen jälkeen. Hallitukset, järjestöt ja yritykset ovat osallistuneet entistä enemmän keskusteluun "oikeudenmukaisesta" ja "eettisestä" tekoälystä ja algoritmeista (Wong 2019).

Teknologian vinoumat voivat johtaa haitallisiin vaikutuksiin, vaikka ne olisivat tahattomia, ja ne aiheuttavat haasteita julkisen luottamuksen rakentamiselle tekoälyyn. Jotta voidaan varmistaa tekoälyn turvallisuus, on otettava huomioon sen erityistekijät ja mahdolliset vaikutukset yhteiskuntaan. Nykyiset toimet tekoälyn ennakoluulojen käsittelemiseksi keskittyvät laskennallisiin näkökohtiin, kuten datan edustavuuteen ja oikeellisuuteen koneoppimisen algoritmeissa. Myös inhimilliset, institutionaaliset ja yhteiskunnalliset tekijät vaikuttavat tekoälyn ennakoluuloihin, ja ne jätetään usein huomiotta. Jotta tähän haasteeseen voidaan vastata onnistuneesti, on meidän laajennettava näkökulmaamme ja tarkasteltava, miten tekoäly muodostuu, ja miten se vaikuttaa yhteiskuntaan.

Luotettava ja vastuullinen tekoäly ei merkitse vain sitä, onko tekoälyjärjestelmä vinoutunut, oikeudenmukainen tai eettinen, vaan myös sitä, tekeekö se todella sitä, mitä se väittää. Läpinäkyvyyden, tietokokonaisuuksien ja TEVV:n (test, evaluation, validation, verification) kaltaiset käytännöt ovat tärkeitä. Myös inhimilliset tekijät, osallistavat suunnittelutekniikat, monien sidosryhmien lähestymistavat ja human-in-the-loop-järjestelmät auttavat pienentämään

tekoälyn vinoumaan liittyviä riskejä. Nämä käytännöt eivät kuitenkaan yksinään tarjoa kattavaa ratkaisua. Tarvitsemme ohjausta laajemmasta sosioteknisestä näkökulmasta, jossa nämä käytännöt yhdistetään yhteiskunnallisiin arvoihin.

Luotettavan ja vastuullisen tekoälyn asiantuntijat suosittelevat näiden arvojen operationalisointia ja uusien normien luomista tekoälyn kehittämistä ja käyttöönottoa varten. Tässä dokumentissa ja CHARLIE-projektin kumppaneiden tekemässä algoritmien ja tekoälyn vinoumaa käsittelevässä työssä omaksutaan sosiotekninen näkökulma.

Vinoumat eivät ole vain tekoälylle ominaisia, ja vinouman riskin poistaminen tekoälyjärjestelmistä on mahdotonta. CHARLIE-hankkeessa on tarkoitus kehittää joukko resursseja ja menetelmiä, joilla voidaan parantaa käytäntöjä vinoumien tunnistamiseksi, ymmärtämiseksi, mittaamiseksi, hallitsemiseksi ja vähentämiseksi. Tämän tavoitteen saavuttamiseksi tarvitaan joustavia tekniikoita, joita voidaan soveltaa eri yhteyksissä, ja joista voidaan tiedottaa eri sidosryhmille.

3. KOMPETENSSIMATRIISI: TEOREETTINEN LÄHESTYMISTAPA

Kompetenssimatriisi, joka tunnetaan myös nimellä osaamismatriisi tai pätevyyskehys, on työkalu, jota käytetään yksilöiden taitojen, tietojen ja kykyjen tunnistamiseen, kartoittamiseen ja arviointiin organisaatiossa tai oppilaitoksessa (Stewart ja Bartrum, 1999). Se tarjoaa visuaalisen esityksen tiettyihin rooleihin, projekteihin, tehtäviin tai akateemisiin ohjelmiin vaadittavasta osaamisesta ja auttaa tunnistamaan työntekijöiden, tiimin jäsenten, opiskelijoiden tai muiden sidosryhmien nykyiset taitotasot (Cottrell, 2018).

Kompetenssimatriisin ominaisuudet

Kompetenssimatriisille on ominaista useat keskeiset piirteet:

- Ruudukko- tai taulukkomuoto: Rivit edustavat yksilöitä ja sarakkeet vaadittuja pätevyyksiä tai taitoja (Mansfield, 2009).
- Luokitukset tai taitotasot: Jokainen matriisin solu sisältää luokituksen tai taitotason vastaavalle taidolle ja yksilölle. Luokitusjärjestelmä voi olla numeerinen, värikoodattu tai perustua kuvaaviin termeihin, kuten "aloittelija", "keskitaso" tai "asiantuntija" (Cottrell, 2018).
- Kustomoitavissa ja mukautettavissa: Matriisit voidaan helposti päivittää vastaamaan rooleissa, vastualueissa tai taitovaatimuksissa tapahtuneita muutoksia (Stewart ja Bartrum, 1999).
- Kattava: Osaamismatriisin tulisi kattaa kaikki tietyssä roolissa, projektissa tai akateemisessa ohjelmassa vaadittavat asiaankuuluvat taidot, ja sen tulisi sisältää sekä tekniset että pehmeät taidot (Mansfield, 2009).

Kompetenssimatriisin päätavoitteet

Kompetenssimatriisilla on useita tärkeitä tavoitteita organisaatiossa tai oppilaitoksessa:

- Taitovajeiden tunnistaminen: Kompetenssimatriisi antaa selkeän yleiskuvan tietyn roolin tai projektin edellyttämistä taidoista ja pätevyyksistä, ja sen avulla voidaan tunnistaa alueet, joilla yksilöt saattavat tarvita lisäkehitystä tai koulutusta (Stewart ja Bartrum, 1999).

- Työntekijöiden tai opiskelijoiden kehityksen tukeminen: Osaamismatriisia voidaan käyttää henkilökohtaisten kehittymissuunnitelmien ohjaamiseen ja koulutus- ja kehittämisaloitteiden tiedottamiseen auttamalla yksilöitä tunnistamaan vahvuutensa ja heikkoutensa sekä asettamaan kehittämisen tavoitteita (Cottrell, 2018).
- Resurssien kohdentamisen optimointi: Kompetenssimatriisi voi auttaa organisaatioita ja oppilaitoksia kohdentamaan resursseja tehokkaammin tunnistamalla tiettyihin rooleihin tai projekteihin sopivimmat henkilöt heidän taitojensa ja osaamisensa perusteella (Mansfield, 2009).
- Suorituskyvyn parantaminen: Kun kompetenssimatriisi auttaa varmistamaan, että yksilöillä on asianmukaiset taidot ja pätevyydet tehtäviinsä, se voi parantaa organisaation tai oppilaitoksen yleistä suorituskykyä (Stewart ja Bartrum, 1999).
- Viestinnän ja yhteistyön helpottaminen: Kompetenssimatriisi voi toimia yhteisenä kielenä, jolla taidoista ja pätevyyksistä keskustellaan organisaatiossa tai oppilaitoksessa, mikä edistää parempaa ymmärrystä ja yhteistyötä tiimin jäsenten, osastojen tai akateemisten yksiköiden välillä (Cottrell, 2018).

Kompetenssimatriisin soveltaminen eri yhteyksissä

Kompetenssimatriiseja on sovellettu eri yhteyksissä, kuten liike-elämässä, julkishallinnossa ja korkeakouluissa, erilaisten tavoitteiden saavuttamiseksi:

- Työvoiman suunnittelussa: Kompetenssimatriisien avulla voidaan tukea strategista henkilöstösuunnittelua tunnistamalla organisaation päämäärien ja tavoitteiden saavuttamiseen tarvittavat taidot ja pätevyydet ja varmistamalla, että ne otetaan huomioon rekrytointi-, koulutus- ja kehittämisaloitteissa (Stewart ja Bartrum, 1999).
- Suorituskyvyn hallinnassa: Matriisit tarjoavat puitteet suorituskykyodotusten asettamiselle, suorituskyvyn arvioinnille ja parannusalueiden tunnistamiselle (Mansfield, 2009).
- Seuraajasuunnittelussa: Kompetenssimatriisit voivat tukea seuraajasuunnittelua yksilöimällä johtotehtävissä vaadittavat taidot ja pätevyydet ja auttamalla tunnistamaan ja kehittämään mahdollisia seuraajia organisaatiossa tai oppilaitoksessa (Cottrell, 2018).

- Opetussuunnitelmien suunnittelussa ja kehittämisessä: Näin varmistetaan, että koulutusohjelmat on suunniteltu siten, että ne kehittävät taitoja ja osaamista, joita opiskelijat tarvitsevat menestyäkseen valitsemillaan aloilla (Billett, 2009). Tämä ja seuraavat kolme muuta kohtaa ovat tämän dokumentin keskeisimmät suuntaviivat.
- Tiedekunnan ja henkilöstön kehittämisessä: Kompetenssimatriiseja voidaan soveltaa korkeakoulujen tiedekunnan ja henkilöstön ammatillisen kehittymisen tukemiseen yksilöllillä taitoja ja valmiuksia, joita tarvitaan tehokkaassa opetuksessa, tutkimuksessa ja hallinnollisissa tehtävissä, ja antamalla tietoa koulutus- ja kehittämisohjelmien suunnittelua varten (Eraut, 2004).
- Tutkimusryhmää muodostettaessa: Osaamismatriiseja voidaan käyttää tutkimusryhmien muodostamisen helpottamiseksi korkeakouluissa tunnistamalla henkilöitä, joilla on toisiaan täydentäviä taitoja ja pätevyys, ja edistämällä monitieteistä yhteistyötä (Börner ym., 2010).
- Opiskelijoiden neuvonnassa ja tukemisessa: Kompetenssimatriiseja voidaan käyttää korkeakoulujen opiskelijoiden neuvonta- ja tukipalvelujen tukemiseen auttamalla neuvoja tunnistamaan opiskelijoiden vahvuudet ja heikkoudet ja ohjaamaan heitä sopivien resurssien, toimenpiteiden tai opiskelupolkujen pariin (Cuseo, 2010).

Kompetenssimatriisin haasteet ja rajoitukset

Vaikka kompetenssimatriisit voivat tarjota arvokasta tietoa ja tukea erilaisia strategisia tavoitteita, niihin liittyy myös joitakin haasteita ja rajoituksia:

- Vaikka kompetenssimatriisit voivat tarjota arvokasta tietoa ja tukea erilaisia strategisia tavoitteita, niihin liittyy myös joitakin haasteita ja rajoituksia:
- Subjektiiivisuus: Osaamisen arviointi voi olla subjektiivista, ja siihen voivat vaikuttaa sellaiset tekijät kuin henkilökohtaiset ennakkoluulot, kulttuurierot tai vaihtelut osaamisen määritelmien tulkinnassa (Stewart ja Bartrum, 1999).
- Aika- ja resurssi-intensiivisyys: Kattavan osaamismatriisin kehittäminen ja ylläpito voi olla aika- ja resurssi-intensiivistä erityisesti suurissa organisaatioissa tai oppilaitoksissa, joissa on erilaisia rooleja ja vastuualueita (Mansfield, 2009).

- Teknisten taitojen ylikorostaminen: Kompetenssimatriiseissa saatetaan painottaa enemmän teknisiä taitoja ja pätevyyskäpepehmeiden taitojen, kuten viestintätaitojen, tiimityöskentelyn tai ongelmanratkaisukyvyn kustannuksella, vaikka nämä taidot ovat usein kriittisiä menestymisen kannalta monissa rooleissa ja yhteyksissä (Cottrell, 2018).
- Jäykkyys: Kompetenssimatriisit voidaan kokea jäykiksi ja joustamattomiksi, mikä saattaa rajoittaa luovuutta, innovointia tai sopeutumiskykyä organisaatiossa tai oppilaitoksessa (Eraut, 2004).

Näistä haasteista ja rajoituksista huolimatta kompetenssimatriisit voivat olla arvokas väline organisaatioille ja oppilaitoksille, jotka haluavat optimoida resurssien jakamista, tukea yksilöiden kehittymistä ja parantaa yleistä suorituskyykyä. Ottamalla järjestelmällisen ja kattavan lähestymistavan taitojen ja pätevyyskäien tunnistamiseen, kartoittamiseen ja arviointiin sekä ottamalla mukaan erilaisia näkökulmia ja sidosryhmiä matriisin kehittämiseen ja ylläpitoon organisaatiot ja oppilaitokset voivat maksimoida tämän lähestymistavan hyödyt ja minimoida mahdolliset haitat (Billett, 2009).

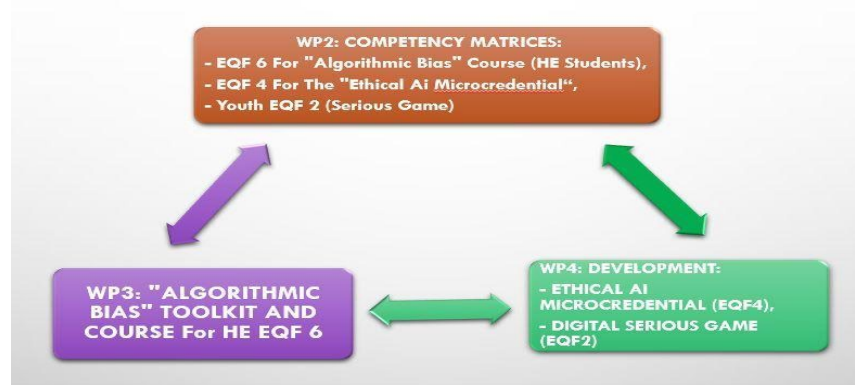
4. CHARLIE TYÖPAKETTI 2 -MATRIISIT

EQF6-kurssin Algoritmin vinouma osaamismatriisi toimii perustana kahdelle muulle kokonaisuudelle: Eettisen tekoälyn mikrokredentiaalikurssi EQF4-osaamismatriisille ja aikuisten ja nuorten hyötypeli EQF2 -oppimistuloksille. Nämä kolme kokonaisuutta liittyvät toisiinsa, ja ne on suunniteltu täydentämään ja vahvistamaan toisiaan, jotta varmistetaan eettisen tekoälyn periaatteiden kattava ymmärtäminen ja soveltaminen eri koulutustasoilla ja -ympäristöissä.

Algoritmin vinouma -kurssin osaamismatriisi luo pohjan Eettinen tekoäly -mikrokredentiaalikurssin EQF4-osaamismatriisin kehittämiseksi. Tarjoamalla vahvan perustan algoritmisten vinoumien ja niiden vaikutusten ymmärtämiselle, tämä kurssi antaa oppijoille mahdollisuuden syventyä tekoälyn eettisiin ulottuvuuksiin EQF4-tasolle edetessään. Eettisen tekoälyn mikrokredentiaalikurssin EQF4-osaamismatriisi rakentuu alustavan osaamisen varaan ja laajentaa sen soveltamisalaa sisältäen laajempia eettisiä näkökohtia ja tekoälyn eettisten periaatteiden käytännön sovelluksia eri aloilla ja skenaarioissa.

Samanaikaisesti Algoritmin vinouma -kurssin osaamismatriisi sisältää myös aikuisten ja nuorten hyötypelin EQF2-oppimistulokset. Näin varmistetaan, että kurssilla kehitetyt peruskäsitteet ja -taidot ovat kiinnostavia ja EQF2-tason oppijoiden saavutettavissa. Hyötypelin muoto on suunniteltu esittelemään nämä aiheet sekä viihdyttävällä että opettavaisella tavalla, ja edistämään algoritmin vinouman ja tekoälyn eettisten periaatteiden syvempää ymmärtämistä laajemman yleisön keskuudessa.

WP2 TO FEED ONTO WPS 3 AND 4:



Kuvio 1. Projektin työpakettien ja tulosten välinen yhteys

Yhteenvedona voidaan todeta, että nämä kolme kokonaisuutta – Algoritmin vinouma -kurssin osaamismatriisi, Eettinen tekoäly -mikrokredentiaalikurssin EQF4-osaamismatriisi ja aikuisten ja nuorten hyötypelin EQF2-oppimistulokset - liittyvät läheisesti toisiinsa, kuten kuvassa on esitetty. Tämä toisiinsa kytkeytyvä rakenne mahdollistaa kattavan ja monitasoisen lähestymistavan algoritmin vinoumasta ja tekoälyn eettisyydestä oppimiseen ja varmistaa, että eri koulutustaustoista ja ikäryhmistä tulevat oppijat voivat kehittää tietoja ja taitoja, joita tarvitaan tekoälyn haasteisiin ja mahdollisuuksiin vastaamiseen eettisesti ja vastuullisesti. Lisäksi WP2:n tulokset vaikuttavat suoraan WP3:een ja WP4:ään, mikä havainnollistaa entisestään näiden osatekijöiden välistä läheistä suhdetta, kuten Kuvio 1 osoittaa.

Tämä työpaketti 2:n, työpaketti 3:n ja työpaketti 4:n saumaton integrointi johtaa vankkaan ja yhtenäiseen opetuskehikseen, joka käsittelee algoritmin vinoumaan ja tekoälyn eettisyyteen liittyviä keskeisiä näkökohtia ja edistää näiden periaatteiden kokonaisvaltaista ymmärtämistä ja soveltamista erilaisissa yhteyksissä ja erilaisissa oppijaprofiileissa.

4.1. ALGORITMIN VINOUMA -KURSSI EQF6

Tämä syventävä koulutusohjelma on suunniteltu tarjoamaan kattava ymmärrys algoritmista vinoumasta ja niiden vaikutuksista nykypäivän datavetoiseen yhteiskuntaan. Koska algoritmeilla on merkittävä rooli elämämme muokkaamisessa ja päätöksenteossa eri aloilla, kuten terveydenhuollossa, koulutuksessa, rahoituksessa ja julkishallinnossa, ammattilaisten ja korkeakouluopiskelijoiden on tärkeää saada vankka käsitys näiden järjestelmien mahdollisista vääristymistä.

Tämän kurssin tavoitteena on antaa osallistujille tiedot ja taidot, joita tarvitaan algoritmien vääristymien tunnistamiseen, lieventämiseen ja käsittelyyn. Teoreettisten perusteiden ja käytännön sovellusten yhdistelmän avulla kurssilaiset saavat käsityksen algoritmiseen vinoumaan liittyvistä eettisistä, sosiaalisista ja teknisistä näkökohdista.

Nämä tavoitteet edistävät yleisiä päämääriä asettamalla kulmakivet seuraaville toimille, jotta opettajilla, ohjaajilla ja oppijoilla olisi selkeä näkemys eri oppimispolkujen tavoitteista ja jotta korkea-asteen koulutuksen, aikuiskoulutuksen ja nuorten koulutuksen valmiuksia tarjota yhteiskunnan tarpeita vastaavia mutta myös oppijoiden oppimistarpeiden mukaan

räätälöityjä oppimismahdollisuuksia voitaisiin lisätä. Työryhmän odotetut tuotokset ovat seuraavat:

- 1) lisätä korkeakoulujen valmiuksia tarjota opiskelijoilleen verkko-oppimismahdollisuuksia, jotka vastaavat yhteiskunnan tarpeisiin, mutta jotka on myös räätälöity opiskelijoiden oppimistarpeiden mukaan - korkeakouluissa hyödynnetään oppimistuloksiin perustuvaa lähestymistapaa, jossa opettajat ja opiskelijat pääsevät yhteisymmärrykseen oppimispolun lopussa odotetuista tuloksista;
- 2) lisätä teknisen alan opiskelijoiden sosiaalisia ja eettisiä taitoja, jotta he voivat suhtautua tekoälyn ja koneoppimisen teknologioihin myönteisesti, kriittisesti ja eettisesti - korkeakouluopiskelijat hyötyvät oppimistuloksiin perustuvasta lähestymistavasta, jossa he tietävät selvästi, mitä heiltä odotetaan oppimispolun päätteeksi;
- 3) antaa opettajille/professoreille digitaalisia ja osallistavia lähestymistapoja aiheen tehokkaaseen opettamiseen (erityisesti verkko-opetuksessa) - sillä oppijakeskeinen lähestymistapa on oppimiskokemusten suunnittelun lähtökohta;
- 4) luoda synergiaa korkea-asteen oppilaitosten sekä alueellisen/paikallisen tason aikuiskoulutuksen ja nuorten opetuksen välille tekoälyn eettisyyden opetuksessa - laatimalla oppimistuloksia toisiaan täydentäville avoimille julkaisuille, jotka on suunnattu eri tasoille (EQF 6, 4, 2) samalla osaamisalalla;
- 5) tehostetaan tekoälyn vinoutumia käsittelevien akateemisten kurssien siirrettävyyttä aikuiskoulutukseen ja nuorten opetukseen - perustamalla yhteyspisteitä eri tuloksissa yhteisen keskustelupohjan luomiseksi oppimispolkujen mahdollisen siirrettävyyden osalta;
- 6) lisätä tietoisuutta aiheesta yhteiskunnallisella tasolla - se on ensimmäinen askel, jonka avulla voidaan tuoda erityisiä avoimia julkaisuja yhteiskunnan eri tasoille ja kohderyhmille.

Kurssin suorittamisen jälkeen osallistujat ymmärtävät kattavasti algoritmin vinoumia, niiden mahdollisia vaikutuksia eri aloihin sekä käytettävissä olevia strategioita ja työkaluja käsitellä näitä vinoumia. Tämä tieto on korvaamatonta ammattilaisille ja korkeakouluopiskelijoille, jotka työskentelevät aloilla, joilla algoritmeja hyödynnetään päätöksenteossa, ja se auttaa varmistamaan tasa-arvoisemmat ja oikeellisemmat tulokset tietoon perustuvassa maailmassa.

4.1.1. Algoritmin vinouma -kurssin sisältö

Kurssi rakentuu viidestä kompetenssiyksiköstä (CU), joista jokaisen tarkoituksena on antaa osallistujille tiedot ja taidot, joita tarvitaan algoritmin vinoumaan liittyvien haasteiden ja mahdollisuuksien kartoittamiseen.

CU1 - Algoritmien mallit ja rajoitukset: Tässä yksikössä esitellään algoritmien peruskäsitteet, niiden mallit ja rajoitukset. Osallistujat saavat käsityksen algoritmien roolista eri aloilla ja niiden käytöstä mahdollisesti aiheutuvista sudenkuopista. Käsiteltäviä aiheita ovat muun muassa algoritmisen kompleksisuus, optimointi ja algoritmisen päätöksenteon rajoitukset.

CU2 - Datan oikeellisuus ja vinoumat tekoälyssä: Yksikössä syvennyttään tekoälyjärjestelmien oikeellisuuden ja vinoumien käsitteisiin ja tutkitaan vinoumien eri lähteitä ja ilmenemismuotoja algoritmin prosessissa. Osallistujat oppivat menetelmiä, joilla voidaan tunnistaa ja mitata datan vääristymiä, sekä strategioita, joilla näihin vääristymiin voidaan puuttua ja niitä voidaan lieventää oikeellisempien tulosten varmistamiseksi.

CU3 - Tekoälyn yksityisyys ja mukautuvuus: Osallistujat tarkastelevat tässä yksikössä yksityisyyden ja mukautuvuuden välistä tasapainoa tekoälysovelluksissa. Keskusteluissa käsitellään haasteita, jotka liittyvät käyttäjien yksityisyyden säilyttämiseen ja samalla yksilöllisten ja tehokkaiden palveluiden tarjoamiseen, sekä olemassa olevia ja kehitteillä olevia tekniikoita, joiden tarkoituksena on tasapainottaa näitä vastakkaisia intressejä.

CU4 - Tekoälyn etiikka, käytännön lähestymistapa: Tämä yksikkö keskittyy eettisten periaatteiden käytännön soveltamiseen tekoälyn kehittämisessä ja käyttöön otossa. Osallistujat tutustuvat erilaisiin eettisiin kehyksiin ja ohjeisiin ja oppivat soveltamaan niitä tekoälyjärjestelmiin liittyviin todellisiin skenaarioihin. Yksikössä käsitellään myös avoimuuden, vastuullisuuden ja sidosryhmien sitoutumisen merkitystä tekoälyn eettisessä kehittämisessä.

CU5 - Tapaustutkimukset ja projekti: Viimeinen yksikkö antaa osallistujille mahdollisuuden soveltaa tietojaan ja taitojaan käytännön tapaustutkimukseen ja käytännön projektiin. Näiden käytännön sovellusten avulla oppijat saavat syvemmän ymmärryksen algoritmin vinouman kompleksisuudesta ja vivahteista sekä käytettävissä olevista strategioista ja välineistä näiden ongelmien ratkaisemiseksi.

4.1.2. Kohderyhmän kuvaus, EQF6-ominaisuudet ja piirteet

Eurooppalainen tutkintojen viitekehys (EQF) on yhteinen viitekehys, joka yhdistää eri Euroopan maiden tutkintojärjestelmät ja helpottaa yksilöiden tutkintojen tunnustamista kaikkialla Euroopassa (Euroopan komissio, 2023). EQF koostuu kahdeksasta viitetasosta, joista kutakin määrittää kuvaaja, joka osoittaa kyseisellä tasolla odotetut oppimistulokset, taidot, pätevyudet ja itsenäisyyden. EQF-taso 6 (EQF6) vastaa tutkintoja, jotka on suoritettu kandidaatin tutkinnon tai ylemmän ammattitutkinnon tasolla Euroopan korkeakoulutusalueella (EHEA). EQF6:n ominaisuuksia ja piirteitä ovat (Euroopan komissio, 2023; CEDEFOP, 2023):

- **Edistynyt tietämys:** EQF6-taso edellyttää, että oppijoilla on tietyn alan syvä tietämys, joka osoittaa kriittistä ymmärrystä teorioista, periaatteista ja menetelmistä. Tällainen tietämys on välttämätöntä omaperäisten ideoiden kehittämiseksi ja monimutkaisten ongelmien ratkaisemiseksi.
- **Kognitiiviset taidot:** EQF6:ssa oppijoiden tulisi kehittää kykyään soveltaa tietojaan uusiin tai tuntemattomiin tilanteisiin, analysoida monimutkaisia ongelmia ja syntetisoida eri lähteistä saatua tietoa. Tähän sisältyvät kriittinen ajattelu, ongelmanratkaisu, päätöksenteko ja luovan ajattelun taidot.
- **Käytännön taidot:** EQF6-tasoon liittyy kehittyneet käytännön taidot, joiden avulla oppijat voivat hallita monimutkaisia teknisiä tai ammatillisia tehtäviä, projekteja tai prosesseja. Näihin taitoihin voivat kuulua tutkimus, projektinhallinta tai kyky käyttää tehokkaasti erikoistyykaluja ja -tekniikoita.
- **Itsenäisyys ja vastuu:** EQF6-oppijoiden odotetaan osoittavan suurta itsenäisyyttä ja vastuullisuutta työssään. Tähän kuuluu vastuun ottaminen omasta oppimisestaan ja ammatillisesta kehittymisestään, muiden työn johtaminen ja valvominen sekä eettisiin ja sosiaalisiin näkökohtiin perustuvien tietoon perustuvien päätösten tekeminen.
- **Viestintä ja yhteistyö:** EQF6-taso edellyttää, että oppijat kehittävät tehokkaita viestintä- ja yhteistyötaitoja, joiden avulla he pystyvät esittämään selkeää ja yksityiskohtaista tietoa, perusteluja tai ehdotuksia sekä asiantuntijoille että muulle yleisölle. Tähän sisältyvät kirjalliset, suulliset ja kommunikaatiotaidot sekä kyky työskennellä tehokkaasti tiimeissä ja eri tieteenalojen välillä.

- **Elinikäinen oppiminen:** EQF6:ssa oppijoiden tulisi kehittää kykyään osallistua elinikäiseen oppimiseen, pohtia omia oppimiskokemuksiaan ja yksilöidä alueita, joilla he voivat kehittyä edelleen. Tähän sisältyy kyky sopeutua uusiin tilanteisiin, teknologioihin tai ammatillisiin ympäristöihin ja osallistua jatkuvaan ammatilliseen kehitykseen.

Seuraavaksi joitakin esimerkkejä EQF6:n tasolle mukautetuista algoritmin vääristymiä koskevista pätevyyksistä:

- **Edistynyt tietämys:** Oppijoilla tulisi olla syvälinen ymmärrys algoritmin vinoumasta, sen lähteistä ja sen vaikutuksesta tekoälyn ja koneoppimisen järjestelmiin. Oppijoiden tulisi tuntea erityyppiset vääristymät, oikeellisuuden mittarit ja vinoutuneiden algoritmien eettiset vaikutukset.
- **Kognitiiviset taidot:** Oppijoiden tulisi pystyä tunnistamaan ja analysoimaan algoritmin vinouman tapauksia reaali maailman tilanteissa, arvioimaan kriittisesti tekoälyn ja koneoppimisen järjestelmien oikeellisuutta ja ehdottamaan menetelmiä vinoutuman lieventämiseksi tai poistamiseksi.
- **Käytännön taidot:** Oppijoiden tulisi kehittää kykyä soveltaa tekniikoita algoritmin vinouman havaitsemiseksi ja lieventämiseksi, kuten uudelleennäytteenottoa, uudelleennäytteenottoa tai vastakohtaista harjoittelua. Oppijoiden tulisi myös pystyä arvioimaan näiden tekniikoiden tehokkuutta vinoumien vähentämisessä ja oikeellisuuden parantamisessa.
- **Itsenäisyys ja vastuu:** Oppijoiden tulisi pystyä tekemään tietoon perustuvia päätöksiä tekoälyn ja koneoppimisen järjestelmien eettisestä käytöstä ottaen huomioon algoritmien vinoutumiin liittyvät mahdolliset riskit ja hyödyt. Oppijoiden tulisi myös pystyä ottamaan vastuuta päätöstensä vaikutuksista yksilöihin, yhteisöihin ja koko yhteiskuntaan.
- **Viestintä ja yhteistyö:** Oppijoiden tulisi pystyä viestimään ymmärryksestään algoritmin vinoumasta ja sen vaikutuksista sekä tekniselle että muulle yleisölle. Oppijoiden tulisi myös pystyä tekemään tehokasta yhteistyötä monitieteisten tiimien kanssa käsitelläkseen tekoälyn ja koneoppimisen oikeudenmukaisuuteen, avoimuuteen ja vastuullisuuteen liittyviä kysymyksiä.

- **Elinikäinen oppiminen:** Oppijoiden tulisi olla sitoutuneita pysymään ajan tasalla uusimmasta tutkimuksesta, menetelmistä ja työkaluista, jotka liittyvät algoritmin vinoumaan, oikeellisuuteen ja eettisyyteen tekoälyyn ja koneoppimiseen liittyen. Heidän tulisi pystyä mukauttamaan tietojaan ja taitojaan vastaamaan alan uusiin haasteisiin ja mahdollisuuksiin.

Yhteenvedona voidaan todeta, että EQF6-tasolle ominaista ovat edistynyt osaaminen, kognitiiviset ja käytännön taidot, itsenäisyys ja vastuullisuus, tehokas viestintä ja yhteistyö sekä sitoutuminen elinikäiseen oppimiseen. Näillä ominaisuuksilla ja piirteillä varmistetaan, että tämän tason tutkinnon suorittaneilla on hyvät valmiudet menestyä valitsemassaan ammatissa tai jatkaa opintojaan korkeammilla tasoilla, kuten suorittamalla maisterin (EQF7) tai tohtorin (EQF8) tutkinnon.

4.1.3. Algoritmin vinouma -kurssin yleinen kompetenssimatriisi EQF6

OPPIMISTULOKSET		
	Valmennuksen suoritettuaan oppija osaa seuraavat asiat	
TIEDOT	TAIDOT	ASENTEET
1. Yhdistää tekoälyn monitieteinen luonne ja algoritmien rooli yhteiskunnan, talouden ja teknologian eri osa-alueiden muotoutumisessa. 2. Saa kattavan käsityksen tekoälyjärjestelmien eettisistä, juridisista ja sosiaalisista vaikutuksista, mukaan lukien oikeudenmukaisuus, yksityisyys ja käytettävyyt. 3. Kehittää kriittisen ajatustavan ja ongelmanratkaisutaitoja, jotta voidaan tunnistaa, analysoida ja käsitellä algoritmien vinoumia ja muita tekoälyyn liittyviä eettisiä haasteita. 4. Arvioida eri lähestymistapojen tehokkuutta algoritmien vinoumien lieventämiseksi ja tekoälyjärjestelmien oikeudenmukaisuuden edistämiseksi ottaen huomioon rajoitukset ja kompromissit. 5. Ymmärtää käyttäjälähtöisen lähestymistavan merkityksen suunniteltaessa ja otettaessa käyttöön	1. Analysoi ja arvioi tekoälyalgoritmeja ja -malleja mahdollisten vinoumien ja rajoitusten tunnistamiseksi ja ehdottaa tarvittaessa parannuksia tai vaihtoehtoja. 2. Soveltaa erilaisia menetelmiä ja tekniikoita datan ja tekoälyjärjestelmien vääristymien tunnistamiseksi, mittaamiseksi ja lieventämiseksi varmistaen oikeudenmukaisuuden ja tasapuoliset tulokset. 3. Tasapainottaa tekoälysovelluksissa yksityisyyden ja käytettävyyden välisiä kompromisseja ja tehdä tietoon perustuvia päätöksiä käyttäjien yksityisyyden suojaamiseksi samalla kun tarjotaan yksilöllisiä ja tehokkaita palveluja. 4. Toteuttaa eettisiä periaatteita, kehäksiä ja ohjeita tekoälyjärjestelmien suunnittelussa, kehittämisessä ja käyttöönotossa ja varmistaa avoimuuden, vastuullisuuden ja	1. Tunnistaa itsenäisesti tekoälyjärjestelmiin liittyvät eettiset kysymykset ja ennakkoluulot ja puuttuu niihin sekä ottaa vastuuta oikeudenmukaisuudesta, yksityisyyden suojasta ja käytettävyydestä kehittämisessä ja käyttöönotossa. 2. Osoittaa sitoutumisensa eettiseen päätöksentekoon tekoälyyn liittyvissä projekteissa, tiedostaa toimiansa mahdolliset seuraukset ja ottaa vastuun valinnoistaan. 3. Pysyy oma-aloitteisesti ajan tasalla nykyisistä ja uusista tekoälyn eettisistä aiheista, säädöksistä ja parhaista käytännöistä ja etsii aktiivisesti mahdollisuuksia jatkuvaan oppimiseen ja ammatilliseen kehittymiseen. 4. Tehokas yhteistyö monialaisten tiimien kanssa, tiedon ja asiantuntemuksen tuottaminen muiden näkökulmia kunnioittaen ja arvostaen sekä vastuun jakaminen projektin tuloksista. 5. Osoittaa vastuuntuntoa yhteiskuntaa kohtaan kehittämällä ja edistämällä tekoälyjärjestelmiä, jotka vastaavat yhteiskunnallisiin tarpeisiin, minimoivat haitat ja

tekoälyjärjestelmiä, jotka kunnioittavat yksityisyyttä, itsenäisyyttä ja yksilöllisiä preferenssejä. 6. Tunnistaa tieteidenvälisen yhteistyön ja sidosryhmien sitoutumisen merkityksen eettisten tekoälyratkaisujen kehittämisessä ja toteuttamisessa. 7. Kehittää vahvan perustan tekoälyyn sovellettaville eettisille teorioille ja viitekehyksille ja oppii soveltamaan näitä periaatteita reaali maailman skenaarioissa. 8. Oppii ymmärtämään avoimuuden, selitettävyyden ja vastuullisuuden merkityksen tekoälyjärjestelmien kehittämisessä ja käyttöönotossa. 9. Ymmärtää sääntelyn, alan standardien ja parhaiden käytäntöjen roolin tekoälyn eettisen kehittämisen ja käyttöönoton muovaamisessa. 10. Kehittää tehokkaita viestintätaitoja, jotta voi osallistua asiantunteviin keskusteluihin ja väittelyihin tekoälyn etiikasta, algoritmista puolueellisuudesta ja niihin liittyvistä aiheista erilaisten yleisöjen kanssa.	sidosryhmien sitouttamisen standardien noudattaminen. 5. Arvioi kriittisesti tekoälyjärjestelmiin liittyviä tapaustutkimuksia ja reaali maailman skenaarioita, tunnistaa eettisiä haasteita ja ehdottaa ratkaisuja kurssilla hankittujen tietojen ja taitojen pohjalta. 6. Tekee tehokasta yhteistyötä monitieteisten tiimien kanssa ja osoittaa ymmärtävänsä erilaisia näkökulmia ja asiantuntemusta, joita tarvitaan tekoälyyn liittyvien monimutkaisten eettisten haasteiden ratkaisemiseksi. 7. Soveltaa käyttäjakeskeisen suunnittelun periaatteita ja strategioita kehittääkseen tekoälyjärjestelmiä, jotka kunnioittavat yksityisyyttä, itsenäisyyttä ja yksilöllisiä preferenssejä. 8. Viestii tekoälyn monitahoisesta etiikasta ja algoritmien vääristymistä ja ratkaisuista selkeästi ja tehokkaasti eri yleisöille, sekä teknisille että muille sidosryhmille. 9. Pysyy ajan tasalla tekoälyn etiikkaan liittyvistä nykyisistä ja uusista säädöksistä, alan standardeista ja parhaista käytännöistä ja soveltaa tätä tietoa sääntöjen mukaisten tekoälyjärjestelmien kehittämiseksi. 10. Harjoittaa jatkuvaa oppimista ja pohdintaa, pyrkien parantamaan ja laajentamaan ymmärrystään tekoälyn etiikasta, algoritmien vääristymistä ja muista asiaan liittyvistä aiheista nopeasti kehittyvässä tekoäly-ympäristössä.	maksimoivat hyödyt eri sidosryhmille. 6. Edistää avoimuutta, vastuullisuutta ja sidosryhmien osallistumista tekoälyn kehittämiseen ja käyttöönottoon, ottaa vastuun eettisistä näkökohdista ja pitää yllä avointa viestintää asianomaisten osapuolten kanssa. 7. Soveltaa eettisiä periaatteita ja ohjeita johdonmukaisesti eri tekoälyprojekteissa ja -konteksteissa, osoittaa henkilökohtaista sitoutumista tekoälyn eettiseen kehittämiseen ja ottaa vastuun näiden standardien noudattamisesta. 8. Toimii itsenäisesti yksilöidessään tekoälyjärjestelmien parannus- tai innovointialueita ja tekee aloitteita eettisiin haasteisiin vastaamiseksi ja oikeudenmukaisuuden edistämiseksi. 9. Tunnistaa ja käsittelee henkilökohtaisia rajoituksia tekoälyn etiikkaan liittyvissä tiedoissa ja taidoissa ja ottaa vastuun tuen, palautteen ja oppimismahdollisuuksien etsimisestä tarpeen mukaan. 10. Edistää eettisen tietoisuuden ja vastuullisuuden kulttuuria ammatillisissa ja akateemisissa yhteisöissään jakamalla tietoa, kokemuksia ja parhaita käytäntöjä tekoälyn etiikkaan ja algoritmien vinoumiin liittyen.
--	---	---

4.1.4. Kompetenssimatriisi kullekin Algoritmin vinouma -kurssin osaamisyksikölle EQF6

CUI - Algoritmit, mallit ja rajoitukset EQF6

OPPIMISTULOKSET

Valmennuksen suoritettuaan oppija osaa seuraavat asiat		
TIEDOT	TAIDOT	ASENTEET
Teoreettinen/tosiasiallinen tietämys: CUK1.1. Ymmärtää algoritmien peruskäsitteitä, niiden malleja ja rajoituksia. CUK1.2. Tunnistaa algoritmien rooli eri aloilla ja niiden käytöstä aiheutuvat mahdolliset sudenkuopat. CUK1.3. Ymmärtää algoritmien vinoumien, optimoinnin ja algoritmisen päätöksenteon rajoitukset.	Osaamistavoite 1: Algoritmien analysointi ja arviointi. 1.1. Arvioi erilaisten algoritmimallien vahvuuksia ja heikkouksia sekä niiden soveltuvuutta eri yhteyksiin ja ongelmiin. 1.2. Arvioi kriittisesti algoritmien rajoituksia päätöksentekoprosesseissa ottaen huomioon sellaiset tekijät kuten laskennallinen kompleksisuus, tiedon laatu ja oikeudenmukaisuus. Osaamistavoite 2: Soveltaa algoritmisia ratkaisuja reaali maailman ongelmiin. 2.1. Valitsee sopivia algoritmeja tiettyihin tehtäviin ottaen huomioon niiden soveltuvuuden kyseiseen ongelmaan ja niiden käyttöön liittyvät mahdolliset sudenkuopat. 2.2. Toteuttaa ja hienosäätää algoritmeja suorituskäytön optimoimiseksi ottaen huomioon sellaiset tekijät kuin tehokkuus, tarkkuus ja skaalautuvuus. Osaamistavoite 3: Algoritmien riskien ja ennakkoluulojen lieventäminen. 3.1. Tunnistaa mahdolliset vinoumien, virheiden tai tahattomien seurausten lähteet algoritmisissa päätöksentekoprosesseissa. 3.2. Kehittää ja soveltaa strategioita algoritmien riskien ja vinoumien minimoimiseksi ja varmistaa tarkemmat, oikeudenmukaisemmat ja eettisemmät tulokset eri aloilla.	Asenteellinen oppimistavoite 1: Eettinen tietoisuus algoritmien kehittämisessä ja käyttöönotossa. 1.1. Osoittaa korkeaa vastuuntuntoa algoritmien kehittämisessä ja käyttöönotossa ottaen huomioon mahdolliset vaikutukset yksilöihin, yhteisöihin ja koko yhteiskuntaan. 1.2. Osoittaa sitoutumista oikeudenmukaisuuteen, avoimuuteen ja vastuullisuuteen algoritmien suunnittelussa ja käytössä ja pyrkii minimoimaan mahdolliset vinoumat ja tahattomat seuraukset. Asenteellinen oppimistavoite 2: Kriittinen ajattelu ja reflektiivinen käytäntö. 2.1. Omaksuu kriittisen ajattelutavan käsitellessään algoritmeja, niiden malleja ja rajoituksia, kyseenalaistaa jatkuvasti oletuksia ja harkitsee vaihtoehtoisia näkökulmia. 2.2. Harjoittaa reflektiivää käytäntöä ja arvioiden henkilökohtaisia ja ammatillisia kokemuksia algoritmeista, jotta voidaan tunnistaa parannusalueita ja jatkuvaa oppimista. Asenteellinen oppimistavoite 3: Yhteistyö ja tieteidenvälinen lähestymistapa. 3.1. Arvostaa yhteistyön ja monitieteisten lähestymistapojen arvoa, kun käsitellään monimutkaisia, todellisia ongelmia, joihin liittyy algoritmeja. 3.2. Etsii tilaisuuksia työskennellä erilaisten tiimien kanssa ja oppii eri alojen asiantuntijoilta kehittääkseen monipuolisia, innovatiivisia ratkaisuja, joilla algoritmisiin haasteisiin vastataan kokonaisvaltaisesti.

**DIGCOMP2.2.¹, ENTRECOMP-²
JA GREENCOMP-³ KEHYSTEN
KOMPETENSSIT CUI
LINKITETTYNÄ**

- Tiedostaa, että tekoäly sinänsä ei ole hyvä eikä huono asia. Se, ovatko tekoälyjärjestelmän tulokset yhteiskunnan kannalta myönteisiä vai kielteisiä, riippuu siitä, miten tekoälyjärjestelmä suunnitellaan ja miten sitä käytetään, kuka sitä käyttää ja mihin tarkoituksiin (DIGCOMP2.2).

- Tiedostaa, että hakukoneet, sosiaalinen media ja sisältöalustat käyttävät usein tekoälyalgoritmeja tuottaakseen vastauksia, jotka on mukautettu yksittäiselle käyttäjälle (esim. käyttäjät näkevät edelleen samanlaisia tuloksia tai sisältöä). Tätä kutsutaan usein "personoinniksi" (DIGCOMP2.2).

- Tiedostaa, että tekoälyalgoritmit toimivat tavoilla, joita käyttäjät eivät yleensä näe tai ymmärrä helposti. Tätä kutsutaan usein "mustan laatikon" päätöksenteoksi, koska voi olla mahdotonta jäljittää, miten ja miksi algoritmi tekee tiettyjä ehdotuksia tai ennusteita (DIGCOMP2.2).

- Tiedostaa, että kaikilla EU:n kansalaisilla on oikeus olla joutumatta täysin automatisoidun päätöksenteon kohteeksi (esim. jos automaattinen järjestelmä hylkää luottihakemuksen, asiakkaalla on oikeus pyytää, että henkilö tarkastaa päätöksen uudelleen). (DIGCOMP2.2)

- Tiedostaa, että tekoäly sinänsä ei ole hyvä eikä huono asia. Se, ovatko tekoälyjärjestelmän tulokset yhteiskunnan kannalta myönteisiä vai kielteisiä, riippuu siitä, miten tekoälyjärjestelmä suunnitellaan ja käytetään, kuka sitä käyttää ja mihin tarkoituksiin. (DIGCOMP2.2.)

- Osaa nimetä joitakin esimerkkejä tekoälyjärjestelmistä: tuotesuosituks (esim. verkkokauppasivustoilla), äänentunnistus (esim. virtuaaliavustajat), kuvantunnistus

**DIGCOMP2.2.-, ENTRECOMP- JA
GREENCOMP- KEHYSTEN
KOMPETENSSIT CUI LINKITETTYNÄ**

- Osaa tunnistaa alat, joilla tekoäly voi tuoda hyötyä jokapäiväisen elämän eri osa-alueille. Esimerkiksi terveydenhuollossa tekoäly voi auttaa varhaisessa diagnosoinnissa, kun taas maataloudessa sitä voidaan käyttää tuholaistuhojen havaitsemiseen (DIGCOMP2.2).

- Osaa muotoilla hakukyselyitä, jotta saadaan aikaan haluttu lopputulos, kun ollaan vuorovaikutuksessa keskusteluagenttien tai älykkäiden kaiuttimien (esim. Siri, Alexa, Cortana, Google Assistant) kanssa, esim. tunnistaa, että jotta järjestelmä pystyy vastaamaan halutulla tavalla, kyselyn on oltava yksiselitteinen ja se on esitettävä selkeästi, jotta järjestelmä voi vastata (DIGCOMP2.2).

**DIGCOMP2.2.-, ENTRECOMP- JA
GREENCOMP- KEHYSTEN KOMPETENSSIT
CUI LINKITETTYNÄ**

- Valmius pohtia tekoälyjärjestelmiin liittyviä eettisiä kysymyksiä (esim. missä yhteyksissä, kuten rikollisten tuomitsemisessa, tekoälysuosituksia ei pitäisi käyttää ilman ihmisen väliintuloa) (DIGCOMP2.2.).

- Punnita tekoälypohjaisten hakukoneiden käytön hyötyjä ja haittoja (esim. vaikka ne saattavat auttaa käyttäjiä löytämään halutun tiedon, ne voivat vaarantaa yksityisyyden ja henkilötietojen suojan tai altistaa käyttäjän kaupallisille intresseille) (DIGCOMP2.2.).

- On halukas oppimaan jatkuvasti, kouluttautumaan ja pysymään ajan tasalla tekoälystä (esim. ymmärtämään, miten tekoälyalgoritmit toimivat, ymmärtämään, miten automaattinen päätöksenteko voi olla puolueellista, erottamaan toisistaan realistisen ja epärealistisen tekoälyn ja ymmärtämään eron suppean tekoälyn eli nykypäivän tekoälyn, joka pystyy suorittamaan kapea-alaisia tehtäviä, kuten pelien pelaamista, ja yleistason tekoälyn eli ihmisälyä paremman tekoälyn välillä, joka on edelleen tieteellistä fiktiota) (DIGCOMP2.2.).

¹<https://publications.jrc.ec.europa.eu/repository/handle/JRC128415>

²https://joint-research-centre.ec.europa.eu/entrecomp-entrepreneurship-competence-framework/competence-areas-and-learning-progress_en

³https://joint-research-centre.ec.europa.eu/greencomp-european-sustainability-competence-framework_en

(esim. kasvainten havaitseminen röntgenkuvissa) ja kasvontunnistus (esim. valvontajärjestelmissä). (DIGCOMP2.2.) - Tiedostaa, että tekoäly on jatkuvasti kehittyvä ala, jonka kehitys ja vaikutukset ovat vielä hyvin epäselviä (DIGCOMP2.2).		
--	--	--

CUI OSAAMISTAVOITTEIDEN KUVAUS

1.1. Ymmärtää algoritmien peruskäsitteet, niiden mallit ja rajoitukset.

Algoritmien peruskäsitteiden, niiden mallien ja rajoitusten ymmärtäminen edellyttää syvällistä syventymistä "algoritmien perusrakennuspalikoihin, joiden juuret ovat matemaattisissa ja laskennallisissa teorioissa" (Smith, 2018, s. 45). Tähän sisältyy suorituskyykyyn ja tarkkuuteen vaikuttavan luontaisen rakenteen, toiminnallisuuden ja mahdollisten rajoitusten ymmärtäminen.

Algoritmit: Knuthin (1997) mukaan "algoritmi on rajallinen joukko sääntöjä, jotka antavat tietyn tyyppisen ongelman ratkaisemiseen tarvittavan operaatiosarjan". Tietojenkäsittelytieteessä ja tekoälyn alalla nämä menettelyt ovat perustavanlaatuisia tietojenkäsittelyssä, päätöksenteossa ja automaatiossa.

Esimerkki: Eukleideen algoritmi on merkittävä tekniikka kahden luvun suurimman yhteisen jakajan (GCD) löytämiseksi, ja se demonstroi prosessuaalisen reduktion periaatetta ratkaisun löytämiseksi iteratiivisesti (Johnson, 2003).

Algoritmimallit: Nämä edustavat "abstrakteja käsitteellisiä kehyksiä, joita käytetään algoritmien rakenteen ja käyttäytymisen ymmärtämiseen ja analysointiin" (Doe & Roe, 2020). Keskeistä on aikakompleksisuuden, tilakompleksisuuden ja algoritmin tehokkuuden tarkastelu.

Esimerkki: Siinä ongelma jaetaan pienempiin, hallittaviin osaongelmiin, jotka ratkaistaan itsenäisesti ennen kuin tulokset yhdistetään alkuperäisen ongelman ratkaisemiseksi (Tanenbaum & Austin, 2012).

Algoritmien rajoitukset: (Li ym., 2017): Tällä tarkoitetaan "rajoituksia tai luontaisia heikkouksia, jotka voivat estää algoritmin suorituskyykyyn, tarkkuuden tai sovellettavuuden" (Li ym., 2017). Nämä rajoitukset voivat johtua eri tekijöistä, kuten ongelman monimutkaisuudesta tai tietojen laadusta.

Esimerkki: TSP-ongelma (travelling salesman problem) on esimerkki laskennallisesta vaikeaselkoisuudesta tilanteissa, joissa on suuria tietokokonaisuuksia. Vaikka on olemassa heuristisia menetelmiä approksimaatioita varten, optimaalisen ratkaisun löytäminen on edelleen monimutkainen tehtävä (Cook, 2016).

Tutustumalla algoritmien perusasioihin, niiden malleihin ja rajoituksiin opiskelijat saavat valmiudet käsitteellistää ja rakentaa tehokkaampia ratkaisuja erilaisiin ongelmiin, jotka perustuvat teoriaan ja käytännön sovelluksiin.

1.2 Tunnistaa algoritmien roolin eri aloilla ja niiden käyttöön liittyvät mahdolliset sudenkuopat.

Algoritmien roolin ymmärtäminen eri sektoreilla ja mahdollisten sudenkuoppien tunnistaminen on olennainen osa kokonaisvaltaisen näkemyksen muodostamista algoritmien käyttöönoton vaikutuksista ja vastuista (Williamson, 2018).

Algoritmien rooli eri aloilla: Kuten Parker et al. (2016) korostaa, "algoritmit ovat läpäisseet kaikki alat ja mullistaneet prosesseja ja päätöksentekoa". Seuraavaksi tarkastelemme joitakin aloja, joilla algoritmien vaikutus on erityisen merkittävä:

- a. Rahoitus
- b. Terveystieteet
- c. Sähköinen kaupankäynti
- d. Kuljetus

Algoritmien käytön mahdolliset sudenkuopat: Tämä aiheuttaa huolta "avoimuudesta ja mahdollisesta liiallisesta riippuvuudesta automaatiosta" (O'Neil, 2016).

- a. Vinoumat ja syrjintä
- b. Yksityisyyden suojaan liittyvät kysymykset
- c. Avoimuuden ja vastuun puute
- d. Liiallinen turvautuminen automaatioon

Algoritmien monitahoisen roolin ja mahdollisten vaarojen tunnistaminen antaa oppijoille valmiudet toimia niiden kanssa vastuullisesti ja eettisesti, mikä edistää kriittistä ajattelua ja eettistä vastuunkantoa digitaali-ikäinä.

1.3 Ymmärtää algoritmien monimutkaisuuden, optimoinnin ja algoritmisen päätöksenteon rajoitukset.

Algoritmien monimutkaisuuden, optimointistrategioiden ja algoritmisten päätöksentekoprosessien rajoitusten ymmärtämisen syventäminen voi auttaa ymmärtämään paremmin algoritmisen vivahteita ja haasteita (Zhang, 2019).

Algoritmien monimutkaisuus: Tämä käsite on pohjimmiltaan algoritmin vaatimien laskennallisten resurssien representaatio, joka ilmaistaan tyypillisesti Big O -merkintätavalla, joka toimii "matemaattisena merkintätapana, joka kuvaa funktion rajoittavaa käyttäytymistä" (Cormen et al., 2009).

Esimerkki: Lineaarisen ja logaritmisen aikakompleksisuuden erottaminen toisistaan antaa mahdollisuuden tarkastella algoritmien, kuten lineaarisen haun ja binäärihaun, tehokkuutta ja tehottomuutta (Sedgewick & Wayne, 2011).

Optimointi: Kuten Karp (2010) toteaa, "optimointi on algoritmin tehostamisen taito", joka käsittää strategiat monimutkaisuuden minimoimiseksi ja suorituskyvyn parantamiseksi erilaisten innovatiivisten lähestymistapojen avulla.

Esimerkki: Tämä havainnollistaa optimoinnin muutospotentialiaalia laskentaprosesseissa (Bentley, 1999).

Algoritmisen päätöksenteon rajoitukset: Ne liittyvät usein "tietojen laatuun ja eettisiin näkökohtiin, jotka joskus johtavat tahattomiin seurauksiin" (Diaz & Sonsini, 2016).

Esimerkki: Rikosoikeuden alalla algoritmiset ennusteet ovat herättäneet keskustelua oikeudenmukaisuudesta ja mahdollisista ennakkoluuloista, mikä korostaa algoritmien käyttöönoton monimutkaista eettistä tilannetta (Angwin et al., 2016).

Algoritmien monimutkaisuuden, optimointistrategioiden ja algoritmisen päätöksenteon luontaisten rajoitusten syvällisen tutkimisen kautta oppijat ovat valmiita liikkumaan algoritmien suunnittelun ja käyttöönoton monimutkaisella kentällä tietoisella ja kriittisellä näkökulmalla.

CU2 - Tietojen oikeudenmukaisuus ja vääristymät tekoälyssä EQF6

OPPIMISTAVOITTEET

	Valmennuksen suoritettuaan oppija osaa seuraavat asiat	
TIEDOT	TAIDOT	ASENTEET

**Teoreettinen/tosiasiallinen
tietämys seuraavilla aloilla:**

CUK2.1. Ymmärtää oikeudenmukaisuuden ja vinoutumien käsitteet tekoälyjärjestelmissä.

CUK2.2. Tunnistaa ennakkoluulojen eri lähteet ja ilmenemismuodot algoritmisissa prosesseissa.

CUK2.3. Oppii menetelmiä, joilla voidaan tunnistaa ja mitata vinoumia datassa.

CUK2.4. Ymmärtää strategioita, joiden avulla voidaan puuttua vinoutumiin ja lieventää niitä tasapuolisempien tulosten varmistamiseksi.

Osaamistavoite 1: Oikeudenmukaisuuden ja ennakkoluulojen käsitteiden ymmärtäminen tekoälyjärjestelmissä.

Taito 1.1: Soveltaa oikeudenmukaisuuden ja puolueellisuuden käsitteitä arvioidakseen tekoälyn perustuvien ratkaisujen eettisiä vaikutuksia eri yhteyksissä.

Taito 1.2: Oikeudenmukaisuuden ja vääristymien huomioon ottaminen tekoälyjärjestelmien suunnittelussa ja kehittämisessä siten, että eettinen ja sosiaalinen vastuu asetetaan etusijalle.

Osaamistavoite 2: Tunnistaa algoritmisissa prosesseissa ilmenevien vinoumien eri lähteet ja ilmenemismuodot:

Taito 2.1: Käyttää analyyttisiä tekniikoita algoritmisissa prosesseissa esiintyvien erityyppisten ennakkoluulojen, kuten dataan perustuvan ennakkoluulon, malliin perustuvan ennakkoluulon ja ihmislähtöisen ennakkoluulon, havaitsemiseksi ja diagnosoimiseksi.

Taito 2.2: Arvioi mahdollisia vääristymien lähteitä tekoälyn kehittämisprosessin eri vaiheissa tietojen keräämisestä mallin käyttöönottoon.

Osaamistulos 3: Oppii menetelmiä, joilla voidaan tunnistaa ja mitata datassa esiintyviä vääristymiä:

Taito 3.1: Käyttää erilaisia menetelmiä ja välineitä, joilla voidaan tunnistaa, mitata ja kvantifioida vinoumia tietokokonaisuuksissa ja algoritmisissa tuotoksissa.

Taito 3.2: Seuraa ja arvioi tekoälyjärjestelmiä jatkuvasti mahdollisten vääristymien ja tahattomien seurausten varalta ja mukauttaa strategioita niiden lieventämiseksi tarpeen mukaan.

Osaamistavoite 4: Ymmärtää strategioita, joilla vääristymiin voidaan puuttua ja niitä voidaan lieventää tasapuolisempien tulosten varmistamiseksi:

Taito 4.1: Kehittää ja toteuttaa strategioita tekoälyjärjestelmien vinoumien käsittelemiseksi ja lieventämiseksi ottaen huomioon kompromissit tarkkuuden, oikeudenmukaisuuden ja muiden suorituskykymittareiden välillä.

Taito 4.2: Arvioi vinoumien lieventämiseen käytettävien tekniikoiden tehokkuutta tekoälysovellusten tasapuolisempien tulosten saavuttamisessa.

Taito 4.3: Tekee yhteistyötä erilaisten sidosryhmien, kuten datatieteilijöiden, insinöörien ja alan asiantuntijoiden, kanssa tekoälyjärjestelmien vastuullisen ja eettisen käytön varmistamiseksi.

Taito 4.4: Viestii tekoälyjärjestelmien oikeudenmukaisuuteen ja vinoumiin liittyvistä haasteista ja ratkaisuista sekä tekniselle että muulle yleisölle.

Asenteellinen oppimistavoite 1: Ymmärtää tekoälyjärjestelmien oikeudenmukaisuuden ja vinoumien käsitteet:

Asenne 1.1: Arvostaa oikeudenmukaisuuden ja tasapuolisuuden merkitystä tekoälyjärjestelmissä ja edistää eettisiä näkökohtia niiden kehittämisessä ja käyttöönotossa.

Asenteellinen oppimistavoite 2: Tunnistaa algoritmisissa prosesseissa esiintyvien vinoumien eri lähteet ja ilmenemismuodot:

Asenne 2.1: Osoittaa ennakoivaa lähestymistapaa mahdollisten vinoumien tunnistamiseen ja käsittelyyn tekoälyn kehitysprosessin aikana.

Asenteellinen oppimistavoite 3: Oppii menetelmiä, joilla voidaan tunnistaa ja mitata datassa esiintyviä vinoumia:

Asenne 3.1: Ymmärtää tarkan ja jatkuvan vinoumien tunnistamisen ja mittaamisen merkityksen tekoälyjärjestelmien vastuullisen kehittämisen ja käytön varmistamisessa.

Asenteellinen oppimistavoite 4: Ymmärtää strategioita, joilla voidaan puuttua vinoumiin ja lieventää niitä tasapuolisempien tulosten varmistamiseksi:

Asenne 4.1: Sitoutuminen tekoälyjärjestelmien vääristymien käsittelemiseen ja lieventämiseen, jotta edistetään tasapuolisia tuloksia kaikille käyttäjille.

Asenne 4.2: Edistää yhteistyön ja jatkuvan oppimisen kulttuuria ja toimii yhteistyössä eri sidosryhmien kanssa tekoälyjärjestelmien vastuullisen ja eettisen käytön varmistamiseksi.

Asenne 4.3: Tunnistaa, että on tärkeää pysyä ajan tasalla tekoälyn etiikkaa, oikeudenmukaisuutta ja vinoumien lieventämistä koskevasta uusimmasta tutkimuksesta, suuntauksista ja parhaista käytännöistä, kehittääkseen jatkuvasti ammatillista pätevyyttään.

	Taito 4.5: Pysyy ajan tasalla tekoälyn etiikan, oikeudenmukaisuuden ja vinoumien lieventämisen alan uusimmasta tutkimuksesta, suuntauksista ja parhaista käytännöistä parantaakseen jatkuvasti ammatillista osaamistaan.	
DIGCOMP2.2.-, ENTRECOMP- JA GREENCOMP- KEHYSTEN KOMPETENSSIT CU2 LINKITETTYNÄ - Tiedostaa eri tekijöiden (esim. tietojen) aiheuttamat mahdolliset tietovääristymät, algoritmit, toimitukselliset valinnat, sensuuri, omat henkilökohtaiset rajoitukset). (DIGCOMP2.2.) - Tietoinen siitä, että tekoälyalgoritmeja ei välttämättä ole konfiguroitu tarjoamaan vain käyttäjän haluamaa tietoa, vaan ne voivat myös sisältää kaupallisen tai poliittisen viestin (esim. kannustaa käyttäjiä pysymään sivustolla, katsomaan tai ostamaan jotakin tiettyä asiaa, jakamaan tiettyjä mielipiteitä). Tällä voi olla myös kielteisiä seurauksia (esim. stereotyyppien toistaminen, väärän tiedon jakaminen). (DIGCOMP2.2.) - Tunnistaa, että vaikka tekoälyjärjestelmien soveltaminen monilla aloilla ei yleensä ole kiistanalaista (esim. tekoäly, joka auttaa torjumaan ilmastonmuutosta), tekoäly, joka on suoraan vuorovaikutuksessa ihmisten kanssa ja tekee heidän elämänsä koskevia päätöksiä, voi usein olla kiistanalaista (esim. CV:n lajitteluohjelmisto rekrytointimenettelyjä varten, kokeiden pisteytys, joka voi määrittää pääsyn koulutukseen) (DIGCOMP2.2.). - Tietoinen siitä, että tekoälyjärjestelmät keräävät ja käsittelevät monenlaisia	DIGCOMP2.2.-, ENTRECOMP- JA GREENCOMP- KEHYSTEN KOMPETENSSIT CU2 LINKITETTYNÄ - Kykenee tunnistamaan, että jotkin tekoälyalgoritmit voivat vahvistaa olemassa olevia näkemyksiä digitaalisissa ympäristöissä luomalla "kaikuja" tai "suodatinkuplia" (esim. jos sosiaalisessa mediassa suositaan tiettyä poliittista ideologiaa, lisäsuositukset voivat vahvistaa tätä ideologiaa ilman, että se altistuu vastakkaisille argumenteille). (DIGCOMP2.2.) - Osaa muuttaa käyttäjän konfiguraatioita (esim. sovelluksissa, ohjelmistoissa, digitaalisissa alustoissa), jotta tekoälyjärjestelmä voi sallia, estää tai kohtuullistaa tietojen seurannan, keräämisen tai analysoinnin (esim. estää matkapuhelinta seuraamasta käyttäjän sijaintia). (DIGCOMP2.2.)	DIGCOMP2.2.-, ENTRECOMP- JA GREENCOMP- KEHYSTEN KOMPETENSSIT CU2 LINKITETTYNÄ - Tunnistaa sekä myönteiset että kielteiset vaikutukset, jotka aiheutuvat siitä, että tekoälyn perustuvassa toiminnassa käytetään kaikkia tietoja (keruu, koodaus ja käsittely), mutta erityisesti henkilötietoja. sovellusten ja verkkopalvelujen kaltaisten digitaalisten teknologioiden avulla. (DIGCOMP2.2.)

käyttäjätietoja (esim. henkilötietoja, käyttäytymistietoja ja kontekstisidonnaisia tietoja) luodakseen käyttäjäprofiileja, joita sitten käytetään esimerkiksi ennustamaan, mitä käyttäjä saattaa haluta nähdä tai tehdä seuraavaksi (esim. tarjota mainoksia, suosituksia, palveluja). (DIGCOMP2.2.) - Tiedostaa, että eettiset käsitteet ja oikeudenmukaisuus nykyisiä ja tulevia sukupolvia kohtaan liittyvät luonnon suojeluun (GREENCOMP).		
--	--	--

CU2 OSAAMISTAVOITTEIDEN KUVAUS

Päätavoitteena on antaa EQF6-tason opiskelijoille syvälinen ymmärrys tietojen oikeudenmukaisuuden ja vinoumien kompleksisuudesta tekoälyssä ja edistää eettisiä ja osallistavia käytäntöjä.

2.1. Oikeudenmukaisuuden ja vinouman käsitteiden ymmärtäminen tekoälyjärjestelmissä

Oppijat kehittävät monipuolisen käsityksen tekoälyjärjestelmien oikeudenmukaisuuden ja vinoumien teoreettisista lähtökohdista. Oppijat tutkivat perustavanlaatuisiin teorioihin nojautuen tekoälyteknologiaan liittyviä moraalisia ja eettisiä vaikutuksia. Kuten Mittelstadt et al. (2016) toteaa, tämä tieto on keskeistä "algoritmisen päätöksenteon eettisten vaikutusten tunnistamisessa ja ymmärtämisessä".

Selitys:

Teoreettiset kehykset: Opiskelijat syventyvät teorioihin ja viitekehyksiin, joissa tutkitaan tekoälyn oikeudenmukaisuuden ja vinoumien käsitteitä.

Eettiset vaikutukset: Analysoidaan näiden käsitteiden eettisiä vaikutuksia reaali maailman skenaarioissa ja edistetään kriittistä lähestymistapaa tekoälyjärjestelmien kehittämiseen.

Esimerkki: Yksityiskohtaiset tapaustutkimukset, joissa tutkitaan algoritmien ennakkoluulojen ilmenemismuotoja eri aloilla, kuten terveydenhuollossa, rahoituksessa ja lainvalvonnassa.

2.2. Tunnistetaan algoritmisissa prosesseissa esiintyvien vinoumien eri lähteet ja ilmenemismuodot

Oppijat valmennetaan paikantamaan vinoumien mahdolliset lähteet ja ilmenemismuodot käyttäen Barocasin ym. (2019) suosittelemaa tutkivaa

lähestymistapaa, jossa kehoitetaan tunnistamaan puolueellisuus "tietojen tuottamisesta päättelyyn".

Selitys:

Datan luomiseen liittyvät vinoumat: Oppijat tarkastelevat kriittisesti, miten ennakkoluuloja voi syntyä tietojen tuottamisprosessin aikana, ja oppivat strategioita niiden estämiseksi.

Päätelmiin liittyvät vinoumat: Oppijat tutkivat, miten vinoumat voivat ilmetä algoritmisissa päätelmissä ja millaisia seurauksia niillä voi olla.

Esimerkki: Analyysi reaali maailman skenaarioista, joissa on havaittu ennakkoluuloja algoritmiprosessien eri vaiheissa, ja kriittinen arviointi toimenpiteistä niiden poistamiseksi.

2.3. Menetelmät, joiden avulla voidaan tunnistaa ja mitata tietojen vinoumia

Painotuksena on käytännön taitojen hankkiminen tietokokonaisuuksien harjojen tunnistamiseksi ja mittaamiseksi. Danksin ja Londonin (2017) mukaan kestävien menetelmien hyödyntäminen on ratkaisevan tärkeää "algoritmisten vinoumien, jotka voivat johtaa perusteettomiin eriyettyihin vaikutuksiin", havaitsemisessa ja lieventämisessä.

Selitys:

Analyysitekniikat: Opiskelijat perehtyvät erilaisiin analyysitekniikoihin ja välineisiin, joiden avulla he voivat tehokkaasti tunnistaa ja mitata tietojen vääristymiä.

Kriittinen arviointi: Opiskelijoita kannustetaan arvioimaan kriittisesti nykyisiä algoritmeja ja ehdottamaan parannuksia virheiden lieventämiseksi.

Esimerkki: Käytännön työpajat, joissa opiskelijat saavat käytännön kokemusta erilaisten analyysitekniikoiden soveltamisesta tietokokonaisuuksien vinoumien tunnistamiseksi ja mittaamiseksi.

2.4. Ymmärrys strategioista, joilla vinoumia voidaan käsitellä ja lieventää tasapuolisempien tulosten varmistamiseksi.

Tämä yksikkö edistää ennakoivaa lähestymistapaa ja painottaa strategioita, joilla voidaan käsitellä ja lieventää tekoälyjärjestelmien vinoumia. Kuten Benjamin (2019) toteaa, nämä strategiat ovat välttämättömiä, jotta voidaan varmistaa sellaisten tekoälyteknologioiden luominen, jotka edistävät "oikeudenmukaisempaa, tasapuolisempaa ja tasa-arvoisempaa yhteiskuntaa".

Selitys:

Lieventävät strategiat: Oppijat tutustuvat erilaisiin strategioihin ja tekniikoihin, joita voidaan käyttää tekoälyjärjestelmien ennakkoluulojen käsittelemiseen ja vähentämiseen.

Eettiset lähestymistavat: Tässä yksikössä korostetaan tekoälyn kehittämiseen liittyvien eettisten lähestymistapojen kehittämistä ja osallisuuden ja oikeudenmukaisuuden edistämistä.

Esimerkki: Projektin kehittäminen, jossa oppijat kehittävät ja toteuttavat strategioita tekoälyjärjestelmien vinoumien poistamiseksi ja edistävät tasapuolisempia tuloksia.

Kun oppijat käyvät huolellisesti läpi tekoälyn liittyvien tietojen oikeudenmukaisuuden ja vinoumien kompleksisuutta, he saavat valmiudet edistää vastuullista ja osallistavaa tekoälykehitystä ja omaksuvat tekoälyn urauurtavien asiantuntijoiden roolin.

CU3 - Tekoälyn yksityisyys ja käytettävyys EQF6

OPPIMISTULOKSET

Valmennuksen suoritettuaan oppija osaa seuraavat asiat		
TIEDOT	TAIDOT	ASENTEET
Teoreettinen/tosiasiallinen tietämys: CUK3.1. Tunnistaa yksityisyyden ja käytettävyyden väliset kompromissit tekoälysovelluksissa. CUK3.2. Ymmärtää haasteet, jotka liittyvät käyttäjien yksityisyyden säilyttämiseen samalla kun tarjotaan yksilöllisiä ja tehokkaita palveluja. CUK3.3. Tutustutaan nykyisiin ja kehitteillä oleviin teknologioihin, joiden tarkoituksena on tasapainottaa yksityisyyden suojaa ja käytettävyyttä tekoälyjärjestelmissä.	Osaamistavoite 1: Tunnistaa tekoälysovellusten yksityisyyden ja käytettävyyden väliset kompromissit. 1.1. Analysoi tekoälysovelluksia koskevia tapaustutkimuksia yksityisyyden ja käytettävyyden välisen tasapainon löytämiseksi. 1.2. Arvioi kriittisesti tekoälyjärjestelmien suunnitteluun liittyviä eettisiä näkökohtia yksityisyyden ja käytettävyyden kannalta. 1.3. Kehittää strategioita, joilla minimoidaan tekoälysovellusten mahdolliset yksityisyydensuojaan liittyvät riskit ja säilytetään samalla käytettävyys. Osaamistavoite 2: Ymmärtää haasteet, jotka liittyvät käyttäjien yksityisyyden säilyttämiseen samalla kun tarjotaan yksilöllisiä ja tehokkaita palveluja. 2.1. Tunnistaa tärkeimmät haasteet, joita tekoälyn kehittäjät kohtaavat varmistaessaan käyttäjien yksityisyyden ja tarjotessaan yksilöllisiä palveluja. 2.2. Arvioi tekoälyjärjestelmien mahdolliset riskit ja haavoittuvuudet, jotka voivat vaarantaa käyttäjien yksityisyyden. 2.3. Ehdottaa ratkaisuja yksityisyyden suojaan liittyvien haasteiden lieventämiseksi, mutta samalla säilytetään tekoälysovellusten yksilöllisten palvelujen tehokkuus. Osaamistavoite 3: Tutustuu nykyisiin ja uusiin teknologioihin, joilla pyritään tasapainottamaan yksityisyyden suojaa ja käyttömukavuutta tekoälyjärjestelmissä.	Asenteellinen oppimistavoite 1: Tunnistaa yksityisyyden ja käytettävyyden väliset kompromissit tekoälysovelluksissa. 1.1. Kehittää vastuuntuntoa yksityisyyden suojaan liittyvien näkökohtien huomioon ottamisesta tekoälysovelluksissa. 1.2. Eettisen ajattelutavan kehittäminen yksityisyyden ja käytettävyyden tasapainottamiseksi tekoälyjärjestelmien suunnittelussa. 1.3. Arvostaa käyttäjien luottamuksen merkitystä luotaessa tekoälysovelluksia, jotka kunnioittavat yksityisyyttä. Asenteellinen oppimistavoite 2: Tiedostaa haasteet, jotka liittyvät käyttäjien yksityisyyden säilyttämiseen samalla kun tarjotaan yksilöllisiä ja tehokkaita palveluja. 2.1. Osoittaa empatiaa käyttäjien yksityisyyden suojaa koskevia huolenaiheita kohtaan yksilöllisten tekoälypalvelujen yhteydessä. 2.2. Edistää sitoutumista yksityisyyden suojaa koskeviin haasteisiin tekoälyn kehittämisessä. 2.3. Arvostaa yksityisyyden suojaa koskevien tekniikoiden jatkuvan parantamisen tärkeyttä yksilöllisten tekoälypalvelujen osalta. Asenteellinen oppimistavoite 3: Tutustuu nykyisiin ja uusiin teknologioihin, joiden tarkoituksena on tasapainottaa yksityisyyden suojaa ja käytettävyyttä tekoälyjärjestelmissä. 3.1. Osoittaa uteliaisuutta ja avoimuutta oppia uusista tekniikoista ja menetelmistä

	3.1. Tutkii ja analysoi nykyisiä teknologioita ja niiden lähestymistapoja yksityisyyden suojaan ja mukavuuteen liittyvien ongelmien ratkaisemiseksi tekoälysovelluksissa. 3.2. Arvioi uusien teknologioiden vahvuuksia ja heikkouksia yksityisyyden suojan ja käytettävyyden tasapainottamisessa tekoälyjärjestelmissä. 3.3. Suunnittelee ja ehdottaa uusia ratkaisuja tai parannuksia olemassa oleviin teknologioihin yksityisyyden ja mukavuuden välisen tasapainon parantamiseksi tekoälysovelluksissa.	yksityisyyden suojaamiseen tähtäävissä tekoälyjärjestelmissä. 3.2. Arvostaa tekoälyn innovatiivisuutta ja sen mahdollisuuksia parantaa yksityisyyttä ja käytettävyyttä. 3.3. Ongelmanratkaisussa on omaksuttava yhteistyöhön perustuva lähestymistapa ja tunnustettava, että tekoälyn liittyvien yksityisyyteen ja mukavuuteen liittyvien huolenaiheiden ratkaiseminen on sidosryhmien yhteinen vastuu.
DIGCOMP2.2.-, ENTRECOMP- JA GREENCOMP- KEHYSTEN KOMPETENSSIT CU3 LINKITETTYNÄ - Tiedostaa, että monissa digitaalisissa teknologioissa ja sovelluksissa (esim. kasvojen seurantakamerat, virtuaaliavustajat, puettavat teknologiat, matkapuhelimet, älylaitteet) käytettävät anturit tuottavat suuria määriä tietoa, myös henkilötietoja, joita voidaan käyttää tekoälyjärjestelmän kouluttamiseen. (DIGCOMP2.2.) - Tiedostaa, että esimerkiksi verkkojärjestelmien keräämää ja käsittelemää dataa voidaan käyttää tunnistamaan kuvioita (esim. toistoja) uudessa datassa (eli muissa kuvissa, äänissä, hiiren napsautuksissa, verkkokäyttäytymisessä), jotta verkkopalveluja (esim. mainoksia) voidaan edelleen optimoida ja personoida. (DIGCOMP2.2.) - Tiedostaa, että kaikkea, mitä ihminen jakaa julkisesti verkossa (esim. kuvat, videot, äänet), voidaan käyttää tekoälyjärjestelmien kouluttamiseen. Esimerkiksi kaupalliset ohjelmistoyritykset, jotka kehittävät tekoälyn kasvojen tunnistusjärjestelmiä, voivat käyttää verkossa jaettuja henkilökohtaisia kuvia (esim. perhevalokuvia) kouluttaakseen ja parantaakseen ohjelmiston kykyä tunnistaa automaattisesti kyseiset henkilöt muissa kuvissa, mikä ei välttämättä ole toivottavaa (esim. yksityisyyden suojan loukkaaminen) (DIGCOMP2.2.).	DIGCOMP2.2.-, ENTRECOMP- JA GREENCOMP- KEHYSTEN KOMPETENSSIT CU3 LINKITETTYNÄ - Osaa käyttää tietotyökaluja (esim. tietokantoja, tiedonlouhinta- ja analyysiohjelmia), jotka on suunniteltu monimutkaisen tiedon hallintaan ja järjestämiseen päätöksenteon ja ongelmanratkaisun tueksi (DIGCOMP2.2.). - Osaa tunnistaa merkkejä, jotka osoittavat, kommunikoidaanko ihmisen vai tekoälyn perustuvan keskusteluagentin kanssa (esim. käytettäessä teksti- tai äänipohjaisia chatbotteja). (DIGCOMP2.2.) - Osaa sisällyttää tekoälyllä muokattua/manipuloitua digitaalista sisältöä omaan työhönsä (esim. sisällyttää tekoälyä ja manipuloitua digitaalista sisältöä). luotuja melodioita omaan musiikkisävellykseen). Tällainen tekoälyn käyttö voi olla kiistanalaista, koska se herättää kysymyksiä tekoälyn roolista taideteoksissa ja esimerkiksi siitä, kenen pitäisi saada siitä kunnianosoitus. (DIGCOMP2.2.) - Kestävän kehityksen arvojen ilmentäminen: 1.2. Oikeudenmukaisuuden tukeminen - nykyisten ja tulevien sukupolvien tasapuolisuuden ja oikeudenmukaisuuden tukeminen ja aiemmilta sukupolvilta oppiminen kestäväyyden edistämiseksi. 1.3 Luonnon vaaliminen - tunnustetaan, että ihminen on osa luontoa, ja kunnioitetaan muiden lajien ja itse luonnon tarpeita ja oikeuksia terveiden ja kestävien ekosysteemien palauttamiseksi ja uudistamiseksi. (GreenComp).	DIGCOMP2.2.-, ENTRECOMP- JA GREENCOMP- KEHYSTEN KOMPETENSSIT CU3 LINKITETTYNÄ - Ottaa huomioon avoimuuden tietoja käsitellessään ja esittäessään niiden luotettavuuden varmistamiseksi ja kiinnittää huomiota tietoihin, jotka on ilmaistu taustalla olevin motiivein (esim. epäeettisin perustein, voittoa tavoittelemalla, manipuloidulla) tai harhaanjohtavilla tavoilla (DIGCOMP2.2.). - Suhtautuu avoimesti tekoälyjärjestelmiin, jotka tukevat ihmisiä tekemään tietoon perustuvia päätöksiä tavoitteidensa mukaisesti (esim. käyttäjät päättävät aktiivisesti, noudattavatko he suositusta vai eivät) (DIGCOMP2.2.). - Ottaa eettisyyden (mukaan lukien, mutta ei rajoittuen ihmisen toimijuuteen ja valvontaan, avoimuuteen, syrjimättömyyteen, esteettömyyteen sekä ennakkoluuloihin ja oikeudenmukaisuuteen) yhdeksi peruspilariksi tekoälyjärjestelmiä kehitettäessä tai otettaessa käyttöön. (DIGCOMP2.2.) - Punnitsee hyödyt ja riskit ennen kuin sallii kolmansien osapuolten käsitellä henkilötietoja (esim. tunnistaa, että älypuhelimien ääniavustaja, jota käytetään kommentojen antamiseen robotti-imurille, voi antaa kolmansille osapuolille - yrityksille, hallituksille, tietoverkkorikollisille - pääsyn tietoihin). (DIGCOMP2.2.) - Ottaa huomioon tekoälyjärjestelmien eettiset seuraukset koko niiden elinkaaren ajan: niihin sisältyvät sekä ympäristövaikutukset (digitaalisten laitteiden ja palvelujen tuotannon ympäristövaikutukset) että yhteiskunnalliset vaikutukset, esim. työn alustamaisuus ja algoritmien johtaminen, joka voi loukata työntekijöiden yksityisyyttä tai oikeuksia; halpatyövoiman käyttö kuvien merkitsemiseen tekoälyjärjestelmien kouluttamiseksi. (DIGCOMP2.2.)

- Tietää, että tekoälyjärjestelmiä voidaan käyttää digitaalisen sisällön (esim. tekstit, uutiset, esseet, twiitit, musiikki, kuvat) automaattiseen luomiseen käyttäen lähteenä olemassa olevaa digitaalista sisältöä. Tällaista sisältöä voi olla vaikea erottaa ihmisen luomuksista. (DIGCOMP2.2.) - Tietää, että henkilötietojen käsittelyyn sovelletaan paikallisia säädöksiä, kuten EU:n yleistä tietosuoja-asetusta (GDPR) (esim. virtuaaliavustajan kanssa käytävä ääniyhteys on GDPR:n kannalta henkilötietoa, ja se voi altistaa käyttäjät tietyille tietosuojaan, yksityisyyteen ja turvallisuuteen liittyville riskeille). (DIGCOMP2.2.) - Tiedostaa, että tietyt toiminnot (esim. tekoälyn kouluttaminen ja Bitcoinin kaltaisten kryptovaluuttojen tuottaminen) ovat resurssi-intensiivisiä prosesseja datan ja laskentatehon osalta. Siksi energiankulutus voi olla suurta, millä voi olla myös suuri ympäristövaikutus. (DIGCOMP2.2.) - Tiedostaa, että tekoäly on inhimillisen älykkyyden ja päätöksenteon tuote (eli ihmiset valitsevat, puhdistavat ja koodaavat datan, suunnittelevat algoritmit, kouluttavat mallit sekä kuratoivat ja soveltavat inhimillisiä arvoja tuotoksiin), eikä se näin ollen ole olemassa ihmisestä riippumatta (DIGCOMP2.2.).		
---	--	--

CU3 OSAAMISTAVOITTEIDEN KUVAUS

Päätavoitteena on kehittää EQF6-tason oppijoille syvälinen ymmärrys yksityisyyden suojan ja käytettävyyden kompleksisesta vuorovaikutuksesta tekoälyssä ja edistää kriittistä tietoisuutta ja taitoja, joita tarvitaan vastuulliseen navigointiin ja innovointiin.

3.1. Ymmärtää yksityisyyden ja käytettävyyden kahtiajaon tekoälyjärjestelmissä

Oppijat perehtyvät intensiivisesti yksityisyyden suojan ja käytettävyyden väliseen tasapainoon. Nämä kaksi keskeistä näkökohtaa määräävät tekoälyjärjestelmien hyväksynnän ja integroinnin yhteiskuntaan. Seuraten Zuboffin (2019) diskurssia, joka selventää "valvontakapitalismia", joka usein

varjostaa teknologista kehitystä, oppijat analysoivat tapaustutkimuksia ja erilaisia näkökulmia, jotka käsittelevät tätä kahtiajakoa.

Selitys:

Historiallinen näkökulma: Tekoälyteknologian kehittymisen myötä yksityisyyden ja käytettävyyden mittakaava on kallistunut ajan mittaan.

Filosofiset lähtökohdat: Oppijat tutkivat yksityisyyteen ja käytettävyyteen liittyviä filosofisia keskusteluja digitaaliaikana.

Esimerkki: Analysoidaan skenaarioita, joissa teknologinen kehitys on rikkonut yksityisyyden suojaa koskevia normeja käytettävyyden ja tehokkuuden tavoittelussa.

3.2. Tunnistaa ja analysoi mahdollisia yksityisyyden suojan loukkauksia tekoälyssä

Oppijat kehittävät taitojaan tunnistaa ja analysoida kriittisesti mahdollisia yksityisyyden suojan loukkauksia erilaisissa tekoälyn toteutuksissa. Tämä on linjassa Pasqualen (2015) tutkimuksen kanssa, jossa hän vaatii "mustan laatikon yhteiskuntaa", jossa **algoritmiprosessien** läpinäkymättömyyttä **olisi tutkittava yksityisyyden suojaamiseksi**.

Selitys:

Analyttiset taidot: Analyttisten taitojen kehittäminen mahdollisten yksityisyyden suojan loukkausten tunnistamiseksi tekoälysovelluksissa.

Oikeudelliset näkökulmat: Yksityisyyden suojaa tekoälyn yhteydessä säätelevien oikeudellisten puitteiden perusteellinen tarkastelu.

Esimerkki: Tapaustutkimukset, joissa analysoidaan korkean profiilin yksityisyyden suojan loukkauksia ja tarkastellaan niistä saatuja kokemuksia ja niiden johdosta toteutettuja toimenpiteitä.

3.3. Tutkia teknisiä edistysaskeleita, joilla pyritään löytämään tasapaino yksityisyyden ja mukavuuden välillä

Tämän opintokokonaisuuden tavoitteena on tutustuttaa opiskelijat uusimpiin teknologisiin edistysaskeliin, joilla pyritään säilyttämään sopusointuinen tasapaino yksityisyyden ja mukavuuden välillä. Opiskelijoita kannustetaan hyödyntämään Nissenbaumin (2009) näkemyksiä, sillä hän kannattaa

"yksityisyyden suojaa kontekstissa" keinona ymmärtää tekoälyyn liittyvien yksityisyyden suojaa koskevien huolenaiheiden vivahteikasta luonnetta.

Selitys:

Kehittyvät teknologiat: Yksityisyyden ja käytettävyyden tasapainottamiseen pyrkivien uusien teknologioiden yksityiskohtainen tarkastelu.

Innovatiiviset lähestymistavat: Analyysi innovatiivisista lähestymistavoista, joita eri alat ovat omaksuneet yksityisyyden suojan sisällyttämisessä mukavuudesta tinkimättä.

Esimerkki: Tutkitaan uusia teknologisia ratkaisuja, kuten eriytettyä yksityisyyden suojaa, jonka tavoitteena on tarjota datan hyötyjä ja samalla säilyttää yksityisyys.

3.4. Kehittää strategioita, joilla edistetään yksityisyyden suoja vaarantamatta tekoälyjärjestelmien käyttömukavuutta

Oppijat saavat mahdollisuuden osallistua aktiivisesti alan kehittämiseen, keskittyen sellaisten strategioiden kehittämiseen, jotka edistävät yksityisyyden suoja vaarantamatta tekoälyjärjestelmien käyttömukavuutta. Opiskelijat perehtyvät Cavoukianin (2010) kaltaisten tutkijoiden teoksiin, jotka ehdottivat "Privacy by Design" -käsitettä tulevaisuuteen suuntautuvaksi lähestymistavaksi yksityisyyden suojan sisällyttämiseksi tekoälyjärjestelmien kehittämiseen.

Selitys:

Strateginen kehittäminen: Kannustetaan oppijoita kehittämään strategioita, joissa otetaan huomioon yksityisyyden suojaaminen vaarantamatta tekoälyn tarjoamaa tehokkuutta ja mukavuutta.

Eettiset näkökohdat: Korostetaan eettisiä näkökohtia, jotka ohjaavat yksityisyyden suojaan keskittyvän tekoälyteknologian kehittämistä.

Esimerkki: Oppijat osallistuvat työpajoihin kehittääkseen tekoälyratkaisuja, joiden keskeisenä ominaisuutena on yksityisyyden suojaaminen, mikä edistää vastuullista innovointia.

Tutustumalla tekoälyn yksityisyyden ja mukavuuden monimutkaiseen kokonaisuuteen oppijat saavat valmiudet analysoida kriittisesti, innovoida ja

johtaa sellaisten tekoälyteknologioiden kehittämistä, jotka kunnioittavat yksityisyyden periaatteita ja tarjoavat samalla vertaansa vailla olevaa mukavuutta.

CU4 - Tekoälyn etiikka, käytännön lähestymistapa EQF6

OPPIMISTULOKSET

TIEDOT	Valmennuksen suoritettuaan oppija osaa seuraavat asiat
	TAIDOT ASENTEET
Teoreettinen/tosiasiallinen tietämys: CUK4.1. Tiedostaa tekoälyn kehittämisen ja käyttöönoton eettisistä periaatteista. CUK4.2. Tietoinen erilaisista eettisistä kehyksistä ja ohjeista sekä niiden soveltamisesta tekoälyjärjestelmiä koskeviin reaali maailman skenaarioihin. CUK4.3. Ymmärtää avoimuuden, vastuullisuuden ja sidosryhmien sitoutumisen merkityksen tekoälyn eettisessä kehitysprosessissa.	Osaamistavoite 1: Soveltaa eettisiä periaatteita tekoälyn kehittämisessä ja käyttöönotossa käytännön lähestymistapojen avulla. 1.1. Soveltaa eettisiä periaatteita tekoälyjärjestelmien suunnittelussa, kehittämisessä ja käyttöönotossa algoritmien vinoumien minimoimiseksi. 1.2. Arvioi tekoälyjärjestelmiä eettisten ohjeiden ja periaatteiden perusteella mahdollisten ennakkoluulojen ja eettisten ongelmien tunnistamiseksi. 1.3. Kehittää strategioita tekoälysovellusten eettisten kysymysten käsittelemiseksi, mukaan lukien mahdolliset ristiriidat periaatteiden ja kompromissien välillä. Osaamistavoite 2: Analysoi ja arvioi erilaisia eettisiä kehyksiä ja ohjeita tekoälyjärjestelmiin liittyvissä todellisissa skenaarioissa. 2.1. Analysoi erilaisia tekoälyn ja algoritmiseen vinoumaan liittyviä eettisiä kehyksiä ja ohjeita. 2.2. Soveltaa eettisiä viitekehyksiä reaali maailman tekoälyskenaarioihin arvioidakseen tekoälyjärjestelmien eettisiä ulottuvuuksia. 2.3. Vertailee erilaisia eettisiä viitekehyksiä sen määrittämiseksi, soveltuvatko ne algoritmiin liittyvien vinoumien käsittelyyn tietyissä tekoälysovelluksissa. Osaamistavoite 3: Suunnittelee ja kehittää eettisiä tekoälyjärjestelmiä edistämällä avoimuutta, vastuullisuutta ja sidosryhmien osallistumista. 3.1. Suunnittelee tekoälyjärjestelmiä, jotka edistävät avoimuutta ja antavat käyttäjille ja sidosryhmille mahdollisuuden ymmärtää päätöksentekoprosessia. 3.2. Toteuttaa vastuullisuustoimenpiteitä sen varmistamiseksi, että tekoälyn kehittäjät ja organisaatiot ottavat vastuun tekoälyjärjestelmiensä seurauksista. 3.3. Edistää sidosryhmien osallistumista tekoälyn kehittämiseen, mukaan lukien eri näkökulmista tulevan palautteen Asenteellinen oppimistavoite 1: Eettinen ajattelutapa tekoälyn kehittämisessä ja käyttöönotossa. 1.1. Kehittää vastuuntuntoa eettisten periaatteiden soveltamisesta tekoälyn kehittämisessä algoritmien puolueellisuuden minimoimiseksi. 1.2. Kehittää eettistä ajattelutapaa, jossa arvostetaan oikeudenmukaisuutta, osallisuutta ja oikeudenmukaisuutta tekoälysovelluksissa. 1.3. Ymmärtää, miten tärkeää on käsitellä eettisiä kysymyksiä ennakoivasti tekoälyjärjestelmien tahattomien seurausten ehkäisemiseksi. Asenteellinen oppimistavoite 2: Uteliaisuus ja avoimuus oppia erilaisista eettisistä kehyksistä ja ohjeista tekoälyjärjestelmiä koskevista todellisissa skenaarioissa. 2.1. Osoittaa uteliaisuutta ja avoimuutta oppia erilaisista eettisistä kehyksistä ja niiden merkityksestä tekoälyn kehittämisessä. 2.2. Ymmärtää eettisten ohjeiden merkitystä tekoälyjärjestelmien suunnittelun ja käyttöönoton suunnittelussa. 2.3. Otetaan käyttöön pohdiskeleva lähestymistapa tekoälysovellusten arvioimiseksi ja jalostamiseksi eettisten näkökohtien perusteella. Asenteellinen oppimistavoite 3: Kannustava ja innostunut ajattelutapa, jolla edistetään avoimuutta, vastuullisuutta ja sidosryhmien sitoutumista tekoälyn eettiseen kehittämiseen. 3.1. Edistää sitoutumista tekoälyjärjestelmien avoimuuden edistämiseen, jotta käyttäjät ja sidosryhmät voivat ymmärtää tekoälyä koskevaa päätöksentekoa ja luottaa siihen. 3.2. Arvostaa vastuullisuuden merkitystä tekoälyn kehittämisessä ja tiedostaa, että organisaatioiden on otettava vastuu tekoälyjärjestelmiensä seurauksista. 3.3. Kannustaa sidosryhmien aktiivista sitoutumista ja yhteistyötä tekoälyn kehittämisessä ja arvostaa erilaisia

	pyytäminen ja tekoälyjärjestelmien vaikutuspiiriin kuuluvien eri yhteisöjen tarpeiden huomioon ottaminen.	näkökulmia ja asianomaisten yhteisöjen osallistumista.
DIGCOMP2.2.-, ENTRECOMP- JA GREENCOMP- KEHYSTEN KOMPETENSSIT CU4 LINKITETTYNÄ - Tiedostaa, että tekoäly ei sinänsä ole hyvä eikä huono asia. Se, ovatko tekoälyjärjestelmän tulokset yhteiskunnan kannalta myönteisiä vai kielteisiä, riippuu siitä, miten tekoälyjärjestelmä suunnitellaan ja käytetään, kuka sitä käyttää ja mihin tarkoituksiin. (Tekoäly) (DigiComp 2.2.) - Tietää, että henkilötietojen käsittelyyn sovelletaan paikallisia säännöksiä, kuten EU:n yleistä tietosuoja-asetusta (GDPR) (esim. virtuaaliavustajan kanssa käytävä puheysteys on GDPR:n kannalta henkilötietoa ja voi altistaa käyttäjät tietyille tietosuojaan, yksityisyyteen ja tietoturvaan liittyville riskeille). (AI) (DigiComp 2.2.)	DIGCOMP2.2.-, ENTRECOMP- JA GREENCOMP- KEHYSTEN KOMPETENSSIT CU4 LINKITETTYNÄ - Osaa tunnistaa alat, joilla tekoäly voi tuoda hyötyä jokapäiväisen elämän eri osaluille. Esimerkiksi terveydenhuollossa tekoäly voi auttaa varhaisessa diagnosoinnissa, kun taas maataloudessa sitä voidaan käyttää tuholaistuhojen havaitsemiseen. (Tekoäly). (DigiComp 2.2.) - Työskentely muiden kanssa (esim. yhteistoiminta, tiimityö, verkoston laajentaminen): Tekee yhteistyötä ja verkostoituu muiden kanssa ideoiden kehittämiseksi ja niiden toteuttamiseksi. Ratkaisee ristiriitoja ja kohtaa kilpailutilanteet tarvittaessa myönteisesti. (EntreComp) Kestävän kehityksen arvojen toteuttaminen: 1.2 Oikeudenmukaisuuden tukeminen - nykyisten ja tulevien sukupolvien tasapuolisuuden ja oikeudenmukaisuuden tukeminen sekä aiemmilta sukupolvilta oppiminen kestävän kehityksen edistämiseksi. 1.3 Luonnon vaaliminen - tunnustetaan, että ihminen on osa luontoa, ja kunnioitetaan muiden lajien ja itse luonnon tarpeita ja oikeuksia terveiden ja joustavien ekosysteemien palauttamiseksi ja uudistamiseksi. (GreenComp)	DIGCOMP2.2.-, ENTRECOMP- JA GREENCOMP- KEHYSTEN KOMPETENSSIT CU4 LINKITETTYNÄ - Valmius pohtia tekoälyjärjestelmiin liittyviä eettisiä kysymyksiä (esim. missä yhteyksissä, kuten rikollisten tuomitsemisessa, tekoälysuosituksia ei pitäisi käyttää ilman ihmisen väliintuloa)? (Tekoäly) (DigiComp 2.2.) - Avoin tekoälyjärjestelmille, jotka tukevat ihmisiä tekemään tietoon perustuvia päätöksiä tavoitteidensa mukaisesti (esim. käyttäjät päättävät aktiivisesti, noudattavatko he suositusta vai eivät). (Tekoäly) (DigiComp 2.2.) - Ottaa eettisyyden (mukaan lukien mutta ei rajoittuen ihmisen toimijuuteen ja valvontaan, avoimuuteen, syrjimättömyyteen, saavutettavuuteen sekä ennakkoluuloihin ja oikeudenmukaisuuteen) yhdeksi keskeiseksi pilariksi tekoälyjärjestelmiä kehitettäessä tai käyttöön otettaessa. (AI) (DigiComp 2.2.) - Avoin osallistumaan yhteistyöprosesseihin, joissa suunnitellaan ja luodaan yhdessä tekoälyjärjestelmiin perustuvia uusia tuotteita ja palveluja, joilla tuetaan ja parannetaan kansalaisten osallistumista yhteiskuntaan. (Tekoäly) (DigiComp 2.2.) Luovuus (esim. kierre on utelias ja avoin): Kehittää luovia ja tarkoituksenmukaisia ideoita. Kehittää useita ideoita ja mahdollisuuksia luoda arvoa, mukaan lukien parempia ratkaisuja nykyisiin ja uusiin haasteisiin. Tutkii ja kokeilee innovatiivisia lähestymistapoja. Yhdistää tietoa ja resursseja arvokkaiden vaikutusten aikaansaamiseksi. (EntreComp)

		Muiden aktivoiminen: Inspiroi ja innostaa asianomaisia sidosryhmiä. Hankkii tarvittavan tuen arvokkaiden tulosten saavuttamiseksi. Osoittaa tehokasta viestintää, vakuuttamista, neuvottelua ja johtajuutta. (EntreComp)
--	--	--

CU4 OSAAMISTAVOITTEIDEN KUVAUS

4.1. Tietoisuus tekoälyn kehittämisen ja käyttöönoton eettisistä periaatteista

Sovelletaan eettisiä periaatteita tekoälyjärjestelmien suunnittelussa, kehittämisessä ja käyttöönotossa algoritmisen vinouman minimoimiseksi. Arvioidaan tekoälyjärjestelmiä eettisten ohjeiden ja periaatteiden perusteella mahdollisten vääristymien ja eettisten ongelmien tunnistamiseksi. Kehitetään strategioita tekoälysovellusten eettisten kysymysten käsittelemiseksi, mukaan lukien mahdolliset ristiriidat periaatteiden ja kompromissien välillä.

Tekoälyn eettiset periaatteet: Eettisillä periaatteilla pyritään luomaan perusta, jonka avulla tekoälyjärjestelmät saadaan toimimaan ihmiskunnan, yksilöiden, yhteiskuntien, ympäristön ja ekosysteemien parhaaksi ja estämään haitat. Ne ohjaavat tekoälyjärjestelmien suunnittelua, kehittämistä, käyttöönottoa ja käyttöä tavalla, joka on luotettava ja jossa keskeistä ovat ihmisarvo, kaikkien ihmisten yhdenvertaisuus, ympäristön, biologisen monimuotoisuuden ja ekosysteemien säilyttäminen, kulttuurisen monimuotoisuuden kunnioittaminen ja tietovastuu (UNESCO, 2022). Eettiset periaatteet kuvaavat, mitä odotetaan oikean ja väärän sekä muiden eettisten normien osalta. Tekoälyn eettiset periaatteet viittaavat normatiivisiin rajoitteisiin, jotka koskevat algoritmien käytön "tekemistä" ja "tekemättä jättämistä" yhteiskunnassa (Zhou et al., 2020). Esimerkiksi tekoälyä käsittelevän korkean tason asiantuntijaryhmän (AI HLEG) mukaan tällaisia periaatteita ovat muun muassa: ihmisen autonomian kunnioittaminen, vahingon ehkäiseminen, oikeudenmukaisuus ja selitettävyyys (Euroopan komissio, 2019).

Tekoälyn eettisten periaatteiden käytännön soveltaminen. Kun eettiset periaatteet on tunnistettu, ne olisi muunnettava toimiviksi työkalupaketeiksi ja suuntaviivoiksi, joilla voidaan muokata tekoälyn perustuvaa innovointia ja tukea tekoälyn eettisten periaatteiden käytännön soveltamista. Työkalupaketteja ja ohjeita siitä, miten eettisiä periaatteita voidaan soveltaa tekoälyn suunnittelussa, toteutuksessa ja käyttöönotossa, tarvitaan kipeästi. Eettiset tekoälyalgoritmit, arkkitehtuurit ja käyttöliittymät noudattavat

tekoälyn eettisiä periaatteita. (Zhou ym., 2020.) Monet yksityiset, julkiset ja voittoa tavoittelemattomat organisaatiot ovat tekoälyn eettisiä periaatteita kehittäessään tunnistaneeet hyödyllisiä toimia, jotka voivat auttaa ehdotettujen periaatteiden artikuloinnissa ja toteuttamisessa (esim. UNESCO, 2021).

Esimerkiksi tekoälyä käsittelevän korkean tason asiantuntijaryhmän hahmottelemat periaatteet on muunnettava konkreettiseksi vaatimuksiksi, jotta voidaan saavuttaa luotettava tekoäly. Näitä vaatimuksia voidaan soveltaa tekoälyjärjestelmien elinkaareen osallistuviin eri sidosryhmiin: kehittäjiin, käyttöönottajiin ja loppukäyttäjiiin sekä laajemmin yhteiskuntaan. Eri sidosryhmillä on eri roolit sen varmistamisessa, että vaatimukset täyttyvät: *kehittäjien* olisi pantava vaatimukset täytäntöön ja sovellettava niitä suunnittelu- ja kehitysprosesseihin, *käyttöönottajien* olisi varmistettava, että heidän käyttämänsä järjestelmät ja heidän tarjoamansa tuotteet ja palvelut täyttävät vaatimukset, *loppukäyttäjille ja laajemmalle yhteiskunnalle* olisi tiedotettava näistä vaatimuksista ja niiden olisi voitava vaatia niiden noudattamista. Alla oleva luettelo vaatimuksista ei ole tyhjentävä. Siihen sisältyy systeemisii, yksilöllisiä ja yhteiskunnallisia näkökohtia:

1. Inhimillinen toimijuus ja valvonta (eli perusoikeudet, inhimillinen toimijuus ja inhimillinen valvonta).
 2. Tekninen kestävyys ja turvallisuus (ts. hyökkäyksenkestävyys ja turvallisuus, varasuunnitelma ja yleinen turvallisuus, tarkkuus, luotettavuus ja toistettavuus).
 3. Yksityisyyden suoja ja tiedonhallinta (eli yksityisyyden suojan kunnioittaminen, tietojen laatu ja eheys sekä tietojen saatavuus).
 4. Avoimuus (eli jäljitettävyyys, selitettävyyys ja viestintä).
 5. Monimuotoisuus, syrjimättömyys ja oikeudenmukaisuus (eli epäoikeudenmukaisten ennakkoluulojen välttäminen, esteettömyys ja yleinen suunnittelu sekä sidosryhmien osallistuminen).
 6. Yhteiskunnallinen ja ympäristöllinen hyvinvointi (eli kestävyys ja ympäristöystävällisyys, sosiaaliset vaikutukset, yhteiskunta ja demokratia).
 7. vastuuvollisuus (ts. tarkastettavuus, kielteisten vaikutusten minimointi ja niistä raportointi, kompromissit ja oikeussuojakeinot).
- (Euroopan komissio, 2019.)

Morley ym. (2019) on laatinut typologian yhdistämällä eettiset periaatteet tekoälyn elinkaaren vaiheisiin, jotta voidaan varmistaa, että tekoälyjärjestelmä suunnitellaan, toteutetaan ja otetaan käyttöön eettisellä tavalla. Typologia osoittaa, että jokainen eettinen periaate olisi otettava huomioon tekoälyn

elinkaaren jokaisessa vaiheessa. Typologia tarjoaa lyhyen tilannekuvan siitä, mitä välineitä tekoälyn kehittäjät voivat tällä hetkellä käyttää edistääkseen eettisen tekoälyn etenemistä periaatteista käytäntöön. Morleyn ym. koko typologia löytyy osoitteesta <https://tinyurl.com/AppliedAIEthics>.

4.2. Tietoa erilaisista eettisistä kehyksistä ja ohjeista sekä niiden soveltamisesta tekoälyjärjestelmiin liittyvissä todellisissa skenaarioissa

Tietoa erilaisista eettisistä kehyksistä ja ohjeista sekä niiden soveltamisesta tekoälyjärjestelmiin liittyvissä reaali maailman skenaarioissa. Analysoidaan erilaisia eettisiä kehyksiä ja ohjeita, jotka liittyvät tekoölyyn ja algoritmisiin vinoumiin. Sovelletaan eettisiä viitekehyksiä tekoölyä koskeviin reaali maailman skenaarioihin tekoälyjärjestelmien eettisten ulottuvuuksien arvioimiseksi. Vertaillaan ja vastakkain asetetaan erilaisia eettisiä viitekehyksiä, jotta voidaan määrittää niiden soveltuvuus algoritmisen vinouman käsittelyyn tietyissä tekoölysovelluksissa.

Tekoälyn etiikkaa koskevat puitteet ja suuntaviivat: Eettiset puitteet ja suuntaviivat koostuvat arvoista, periaatteista, standardeista ja toimista, joilla ohjataan yksilöitä, ryhmiä, yhteisöjä, instituutioita ja yksityisen sektorin yrityksiä varmistamaan etiikan sisällyttäminen tekoälyjärjestelmien elinkaaren kaikkiin vaiheisiin. Niillä pyritään suojelemaan, edistämään ja kunnioittamaan ihmisoikeuksia ja perusvapauksia, ihmisarvoa ja tasa-arvoa, turvaamaan nykyisten ja tulevien sukupolvien edut, säilyttämään ympäristö, biologinen monimuotoisuus ja ekosysteemit sekä kunnioittamaan kulttuurista monimuotoisuutta tekoälyjärjestelmän elinkaaren kaikissa vaiheissa (Euroopan komissio, 2019; UNESCO, 2021).

Erilaisia tekoölyä koskevia eettisiä periaatteita ja kehyksiä on julkaistu tyypillisesti teollisuuden (esim. Google, IBM, Microsoft, Intel), hallitusten (esim. UK Lords Select Committee, Euroopan komission korkean tason asiantuntijaryhmä) ja akateemisten tahojen (esim. Future of Life Institute, IEEE, AI4People) toimesta (Zhou ym., 2020). Kaikki olemassa olevat eettiset kehykset ja suuntaviivat eivät nimenomaisesti tue mitään yksittäistä eettistä periaatetta, mutta seuraavien periaatteiden ympärille on syntymässä yhteneväisyyttä: avoimuus, oikeudenmukaisuus ja tasapuolisuus, vastuullisuus, vahingon välttäminen (non-maleficence), yksityisyys, hyvántahtoisuus (beneficence), vapaus ja autonomia, luottamus, kestävyys, ihmisarvo ja solidaarisuus (Jobin et al, 2019).

Unesco (2021) on julkaissut marraskuussa 2021 ensimmäisen maailmanlaajuisen tekoälyn etiikkaa koskevan standardin - suosituksen tekoälyn etiikasta. Kaikki 193 jäsenvaltiota hyväksyivät tämän kehyksen.

Suosituksen kulmakivenä on ihmisoikeuksien ja ihmisarvon suojelu, joka perustuu niiden peruseriaatteiden edistämiseen.

4.3. Ymmärtää avoimuuden, vastuullisuuden ja sidosryhmien sitoutumisen merkityksen tekoälyn eettisessä kehitysprosessissa

Avoimuuden, vastuullisuuden ja sidosryhmien sitoutumisen merkityksen ymmärtäminen tekoälyn eettisessä kehittämisessä. Suunnitellaan tekoälyjärjestelmiä, jotka edistävät läpinäkyvyyttä ja antavat käyttäjille ja sidosryhmille mahdollisuuden ymmärtää päätöksentekoprosessia. Toteutetaan vastuullisuustoimenpiteitä sen varmistamiseksi, että tekoälyn kehittäjät ja organisaatiot ottavat vastuun tekoälyjärjestelmiensä seurauksista. Sidosryhmien sitoutumisen helpottaminen tekoälyn kehittämiseen, mukaan luettuna eri näkökulmista tulevan palautteen pyytäminen ja tekoälyjärjestelmien vaikutuspiiriin kuuluvien yhteisöjen tarpeiden huomioon ottaminen.

Avoimuus: Läpinäkyvyys (selitettävyys) on ratkaisevan tärkeää, jotta käyttäjien luottamus tekoälyjärjestelmiin voidaan rakentaa ja ylläpitää. Tämä tarkoittaa, että prosessien on oltava selkeitä, tekoälyjärjestelmien kyvyistä ja tarkoituksesta on tiedotettava avoimesti ja päätökset ovat mahdollisuuksien mukaan voitava selittää niille, joita asia suoraan tai välillisesti koskee. Ilman tällaista tietoa päätöstä ei voida kyseenalaistaa asianmukaisesti.

Aina ei ole mahdollista selittää, miksi malli on tuottanut tietyn tuotoksen tai päätöksen (ja mikä yhdistelmä syöttötekijöitä on vaikuttanut siihen). Tällaisia tapauksia kutsutaan "mustan laatikon" algoritmeiksi, ja niihin on kiinnitettävä erityistä huomiota. Tällaisissa tapauksissa voidaan vaatia muita selitettävyystoimenpiteitä (esim. jäljitettävyys, tarkastettavuus ja avoin tiedottaminen järjestelmän ominaisuuksista) edellyttäen, että järjestelmä kokonaisuutena kunnioittaa perusoikeuksia. Se, missä määrin selitettävyyttä tarvitaan, riippuu suuresti asiayhteydestä ja seurausten vakavuudesta, jos kyseinen tuotos on virheellinen tai muuten epätarkka (Euroopan komissio, 2019.)

Vastuullisuus: Vastuullisuus on yksi tekoälyn hallinnan kulmakivistä. Tilivelvollisuudella on monia määritelmiä, mutta sen ytimessä on velvollisuus tiedottaa ja perustella oma toiminta viranomaiselle (Novelli, Taddeo & Floridi, 2023). Yleisesti ottaen tekoälyn vastuullisuus liittyy odotukseen, että suunnittelijat, kehittäjät ja käyttöönottajat noudattavat standardeja ja lainsäädäntöä, joilla varmistetaan tekoälyn asianmukainen toiminta sen elinkaaren aikana (Fjeld ym., 2020).

Olisi kehitettävä asianmukaisia valvonta-, vaikutustenarviointi-, tarkastus- ja due diligence -mekanismeja, mukaan lukien ilmiantajien suojeleminen, jotta voidaan varmistaa tekoälyjärjestelmiä ja niiden vaikutuksia koskeva vastuuvastuuvelvollisuus koko niiden elinkaaren ajan. Sekä teknisillä että institutionaalisilla suunnitelmilla olisi varmistettava tekoälyjärjestelmien (toiminnan) tarkastettavuus ja jäljitettävyyden, jotta voidaan puuttua erityisesti mahdollisiin ristiriitoihin ihmisoikeusnormien ja -standardien kanssa sekä ympäristön ja ekosysteemien hyvinvointiin kohdistuviin uhkiin (UNESCO, 2021).

Sidosryhmien osallistuminen: On tärkeää kuulla asiaankuuluvia sidosryhmiä, kun kehitetään ja hallinnoidaan tekoälysovellusten käyttöä (Fjeld ym., 2020). Eri sidosryhmien osallistuminen tekoälyjärjestelmän koko elinkaaren ajan on välttämätöntä, jotta tekoälyn hallinnointiin voidaan soveltaa osallistavia lähestymistapoja, joiden avulla hyödyt voidaan jakaa kaikkien kesken ja edistää kestävä kehitys. Sidosryhmiä ovat muun muassa hallitukset, hallitustenväliset järjestöt, tekninen yhteisö, kansalaisyhteiskunta, tutkijat ja akateemiset tahot, tiedotusvälineet, koulutus, poliittiset päättäjät, yksityisen sektorin yritykset, ihmisoikeusinstituutit ja tasa-arvoelimet, syrjinnän vastaiset valvontaelimet sekä nuoriso- ja lapsiryhmät (UNESCO, 2021).

Sidosryhmien mukaan ottaminen auttaa varmistamaan, että tekoälyteknologiaa kehitetään ja otetaan käyttöön tavalla, joka vastaa teknologian vaikutuspiiriin kuuluvien yksilöiden ja ryhmien arvoja, tarpeita ja huolenaiheita. Käytännössä voi kuitenkin olla vaikeaa tunnistaa ihmisiä ja organisaatioita, joihin tekoälyjärjestelmien kehittäminen ja käyttö vaikuttaa. Nämä kysymykset korostuvat entisestään, kun tekoälyjärjestelmiä kehitetään projekteissa, jotka ovat väliaikaisia organisaatioita, joiden tiimin jäsenet eivät välttämättä ole käytettävissä projektin päätyttyä. Jotkin tekoälyjärjestelmät toimivat automaattisesti eivätkä salli ihmisen puuttumista niihin, mikä tekee niistä ainutlaatuisia verrattuna muihin tietojärjestelmiin. Nämä järjestelmät voivat vaikuttaa ihmis-, kansalais- ja työntekijöiden oikeuksiin; niiden ominaispiirteet ovat ihmisvaikutustensa osalta lähellä lääke- ja terveydenhuoltoalaa ja ympäristövaikutustensa osalta lähellä infrastruktuurihankkeita (Miller, 2022).

CU5 - Tapaustutkimukset ja projekti EQF6

OPPIMISTULOKSET

	Pelillistetyn oppimiskokemuksen jälkeen oppija osaa seuraavat asiat	
TIEDOT	TAIDOT	ASENTEET

Teoreettinen/tosiasiallinen tietämys: CUK5.1. Ymmärtää algoritmin vinouman ja sen seuraukset todellisessa maailmassa. CUK5.2. Tietoisuus algoritmin vinouman kompleksisuudesta ja vivahteista. CUK5.3. Tunnistaa käytettävissä olevat strategiat ja välineet, joilla voidaan puuttua algoritmiseen vinoumaan liittyviin ongelmiin käytännössä.	Osaamistavoite 1: Soveltaa algoritmin vinoumaa koskevia tietoja reaali maailman tapaustutkimuksiin ja käytännön projektiin. 1.1. Tutkitaan reaali maailman tapaustutkimuksia algoritmin puolueellisuuden ja sen seurausten tunnistamiseksi. 1.2. Kehitetään ratkaisuja algoritmin puolueellisuuden käsittelemiseksi tapaustutkimuksissa ottaen huomioon eettiset ja käytännölliset vaikutukset. 1.3. Suunnitellaan, toteutetaan ja arvioidaan käytännönläheinen projekti, jossa käsitellään algoritmin vinoumaa valitulla alalla. Osaamistavoite 2: Analysoi algoritmiajattelua, sen syitä ja vaikutuksia. 2.1. Analysoi erityyppisiä algoritmisia vääristymiä ja niiden perimmäisiä syitä. 2.2. Arvioi algoritmin vinouman mahdollisia vaikutuksia yksilöihin ja koko yhteiskuntaan. 2.3. Ymmärtää algoritmien ennakkoluulojen tunnistamiseen, mittaamiseen ja lieventämiseen liittyvät haasteet ja rajoitukset. Osaamistavoite 3: Tarkastelee ja yhdistää käytettävissä olevia strategioita ja välineitä, joilla voidaan puuttua algoritmiin ennakkoluuloihin. 3.1. Tutkitaan ja arvioidaan olemassa olevia strategioita algoritmin vinouman käsittelemiseksi, mukaan lukien oikeudenmukaisuustietoinen koneoppiminen ja puolueellisuustarkastukset. 3.2. Kehitetään taitoja käyttää välineitä ja tekniikoita tekoälyjärjestelmissä esiintyvien algoritmien vääristymien havaitsemiseksi, mittaamiseksi ja lieventämiseksi. 3.3. Integroidaan monitieteisiä lähestymistapoja, kuten alan asiantuntijoiden panos ja erilaiset näkökulmat, tekoälyjärjestelmien oikeudenmukaisuuden parantamiseksi.	Asenteellinen oppimistavoite 1: Sitoutuminen ongelmanratkaisussa käsittelemään algoritmin vinoumaa ja sen vaikutuksia reaali maailmaan. 1.1. Osoittaa sitoutumisen teoreettisen tiedon soveltamiseen käytännön tilanteisiin algoritmiin liittyvien vinoumien poistamiseksi. 1.2. Arvostaa käytännön kokemuksen merkitystä algoritmiin liittyvien vinoumien ja niiden todellisten vaikutusten syvällisemmän ymmärtämisen kannalta. 1.3. Kehittää ongelmanratkaisukykyä, jolla pyritään luomaan oikeudenmukaisia ja tasapuolisia tekoälyratkaisuja. Asenteellinen oppimistavoite 2: Kriittinen ja yhteistyöhön perustuva ajattelutapa, jolla tunnistetaan algoritmiin liittyvien vinoumien kompleksisuus ja vivahteet. 2.1. Edistää empatiaa niitä yksilöitä ja yhteisöjä kohtaan, joihin algoritmien vinoumat vaikuttavat. 2.2. Kehitetään kriittistä ajattelutapaa, jossa tunnistetaan algoritmin vinoumaan puuttumisen kompleksisuus ja rajoitukset. 2.3. Arvostaa monitieteisen yhteistyön ja erilaisten näkökulmien merkitystä algoritmiseen puolueellisuuteen liittyvien haasteiden ratkaisemisessa. Asenteellinen oppimistavoite 3: Avoin ja yhteistyöhön perustuva ajattelutapa, kun hyödynnetään käytettävissä olevia strategioita ja välineitä algoritmin vinoumaan liittyvien ongelmien ratkaisemiseksi käytännössä. 3.1. Omaksuu ennakoivan lähestymistavan käytettävissä olevien strategioiden ja välineiden hyödyntämiseen algoritmin vinouman torjumiseksi. 3.2. Arvostaa jatkuvan oppimisen merkitystä ja pysyy ajan tasalla algoritmin vinouman käsittelyyn liittyvästä uudesta kehityksestä. 3.3. Kannustetaan yhteistyöhön perustuvaa ympäristöä, joka edistää avointa vuoropuhelua, tiedon jakamista ja innovointia algoritmin vinoumaan liittyvien kysymysten käsittelyssä.
---	---	--

DIGCOMP2.2.-, ENTRECOMP- JA GREENCOMP- KEHYSTEN KOMPETENSSIT CU5 LINKITETTYNÄ

- Tietoinen siitä, että tekoälyalgoritmeja ei välttämättä ole konfiguroitu tarjoamaan vain käyttäjän haluamaa tietoa, vaan ne voivat myös sisältää kaupallisen tai poliittisen viestin (esim. kannustaa käyttäjiä pysymään sivustolla, katsomaan tai ostamaan jotakin tiettyä asiaa tai jakamaan tiettyjä mielipiteitä). Tällä voi olla myös kielteisiä seurauksia (esim. stereotyyppien toistaminen, väärän tiedon jakaminen). (AI) (DigiComp 2.2.)

- Tietoisuus siitä, että tekoälyn perustana oleviin tietoihin voi sisältyä ennakkoluuloja. Jos näin on, tekoälyn käyttö voi automatisoida ja pahentaa näitä ennakkoluuloja. Esimerkiksi ammattia koskevat hakutulokset voivat sisältää stereotyyppioita mies- tai naistyöpaikoista (esim. miespuoliset bussinkuljettajat, naispuoliset myyjät). (Tekoäly) (DigiComp 2.2.)

- Tietää, että esimerkiksi verkkojärjestelmien keräämiä ja käsittelemiä tietoja voidaan käyttää. Tunnistaa kuvioita (esim. toistoja) uusissa tiedoissa (esim. muut kuvat, äänet, hiiren napsautukset), verkkokäyttäytyminen), jotta verkkopalveluja (esim. mainoksia) voidaan edelleen optimoida ja personoida. (DigiComp 2.2.)

- On tietoinen siitä, että monissa digitaalisissa teknologioissa ja sovelluksissa (esim. kasvojenseurantakamerat, virtuaaliavustajat, puettavat teknologiat, matkapuhelimet, älylaitteet) käytettävät anturit tuottavat suuria määriä tietoa, myös henkilötietoja, joita voidaan käyttää tekoälyjärjestelmän kouluttamiseen. (Tekoäly) (DigiComp 2.2.)

- Tietoinen siitä, että kaikkea, mitä ihminen jakaa julkisesti verkossa (esim. kuvat, videot, äänet), voidaan käyttää tekoälyjärjestelmien kouluttamiseen. Esimerkiksi kaupalliset ohjelmistoyritykset, jotka kehittävät tekoälyn

DIGCOMP2.2.-, ENTRECOMP- JA GREENCOMP- KEHYSTEN KOMPETENSSIT CU5 LINKITETTYNÄ

-Eettinen ja kestävä ajattelu (esim. eettinen vuorovaikutus, ovat vastuullisia, arvioivat vaikutuksia): Arvioi ideoiden, mahdollisuuksien ja toimien seurauksia ja vaikutuksia. (...) Toimii vastuullisesti. (EntreComp)

Kestävän kehityksen arvojen toteuttaminen:

1.2 Oikeudenmukaisuuden tukeminen - nykyisten ja tulevien sukupolvien tasapuolisuuden ja oikeudenmukaisuuden tukeminen sekä aiemmilta sukupolvilta oppiminen kestävyuden edistämiseksi.

1.3 Luonnon vaaliminen - tunnustaa, että ihminen on osa luontoa, ja kunnioittaa muiden lajien ja itse luonnon tarpeita ja oikeuksia terveiden ja joustavien ekosysteemien palauttamiseksi ja uudistamiseksi. (GreenComp)

DIGCOMP2.2.-, ENTRECOMP- JA GREENCOMP- KEHYSTEN KOMPETENSSIT CU5 LINKITETTYNÄ

- Avoin tekoälyjärjestelmille, jotka tukevat ihmisiä tekemään tietoon perustuvia päätöksiä tavoitteidensa mukaisesti (esim. käyttäjät päättävät aktiivisesti, noudattavatko he suositusta vai eivät). (Tekoäly) (DigiComp 2.2.)

- Valmius pohtia tekoälyjärjestelmiin liittyviä eettisiä kysymyksiä (esim. missä yhteyksissä, kuten rikollisten tuomitsemisessa, tekoälyn suosituksia ei pitäisi käyttää ilman ihmisen väliintuloa)? (Tekoäly) (DigiComp 2.2.)

- Yksilöidään sekä myönteisiä että kielteisiä vaikutuksia, joita on kaikkien tietojen (keruu, koodaus ja käsittely), mutta erityisesti henkilötietojen, käytöllä tekoälyn perustuvissa digitaalisissa teknologioissa, kuten sovelluksissa ja verkkopalveluissa. (Tekoäly) (DigiComp 2.2.)

- Tarkastellaan tekoälyjärjestelmien eettisiä seurauksia niiden koko elinkaaren aikana: niihin sisältyvät sekä ympäristövaikutukset (digitaalisten laitteiden ja palvelujen tuotannon ympäristövaikutukset) että yhteiskunnalliset vaikutukset, esimerkiksi työn alustamaisuus ja algoritmien johtaminen, joka voi rajoittaa työntekijöiden yksityisyyttä tai oikeuksia; halpatyövoiman käyttö kuvien merkitsemiseen tekoälyjärjestelmien kouluttamiseksi. (Tekoäly) (DigiComp 2.2.)

- Avoin osallistumaan yhteistyöprosesseihin, joissa suunnitellaan ja luodaan yhdessä tekoälyjärjestelmiin perustuvia uusia tuotteita ja palveluja, joilla tuetaan ja parannetaan kansalaisten osallistumista yhteiskuntaan. (Tekoäly) (DigiComp 2.2.)

kasvojentunnistusjärjestelmiä, voivat käyttää verkossa jaettuja henkilökohtaisia kuvia (esim. perhevalokuvia) kouluttaakseen ja parantaakseen ohjelmiston kykyä tunnistaa automaattisesti kyseiset henkilöt muissa kuvissa, mikä ei ehkä ole toivottavaa (esim. yksityisyyden loukkaus). (AI) (DigiComp 2.2.) - Tunnistaa, että vaikka tekoälyjärjestelmien soveltaminen monilla aloilla on yleensä kiistatonta (esim. tekoäly, joka auttaa torjumaan ilmastonmuutosta), tekoäly, joka on suoraan vuorovaikutuksessa ihmisten kanssa, ja heidän elämäänsä koskevat päätökset voivat usein olla kiistanalaisia (esim. CV:n lajitteluohjelmistot rekrytointimenettelyjä varten, kokeiden pisteytys, joka voi määrittää pääsyn koulutukseen). (AI) (DigiComp 2.2.) - Pystyy tunnistamaan joitakin esimerkkejä tekoälyjärjestelmistä: tuotesuosituksia (esim. verkkokauppasivustoilla), äänentunnistus (esim. virtuaaliavustajat), kuvantunnistus (esim. kasvainten havaitseminen röntgenkuvissa) ja kasvontunnistus (esim. valvontajärjestelmissä). (Tekoäly) (DigiComp 2.2.)		
---	--	--

CU5 OSAAMISTAVOITTEIDEN KUVAUS

5.1. Ymmärtää algoritmin vinouma ja sen seuraukset todellisessa maailmassa

Algoritmin vinouma ja sen todellisten vaikutusten ymmärtäminen. Sovelletaan tietoa algoritmista vinoumasta reaali maailman tapaustutkimuksiin ja käytännön projektiin. Tutkitaan reaali maailman tapaustutkimuksia algoritmisen vinouman ja sen seurausten tunnistamiseksi. Kehitetään ratkaisuja, joilla puututaan tapaustutkimuksissa esiintyvään algoritmiseen puolueellisuuteen, ottaen huomioon eettiset ja käytännön seuraukset. Suunnitellaan, toteutetaan ja arvioidaan käytännön projekti algoritmisen ennakoluulon poistamiseksi valitulla alalla.

Algoritmin vinouma: Algoritmin vinoumalla tarkoitetaan epäoikeudenmukaisten tai syrjivien tulosten esiintymistä algoritmisen järjestelmän tuottamissa tuloksissa. Sitä esiintyy, kun algoritmi

johdonmukaisesti ja järjestelmällisesti suosii tai asettaa epäedulliseen asemaan tiettyjä yksilöitä tai ryhmiä tiettyjen ominaisuuksien, kuten rodun, sukupuolen tai sosioekonomisen aseman perusteella. Koneoppimisessa algoritmit perustuvat useisiin tietosarjoihin eli harjoittelutietoihin, joissa määritetään, mitkä ovat oikeat tuotokset joillekin ihmisille tai kohteille. Koulutusdatan perusteella kone oppii mallin, jota voidaan soveltaa muihin ihmisiin tai esineisiin, ja tekee ennusteita niiden mahdollisista tuloksista. Koneet voivat kuitenkin kohdella samankaltaisia ihmisiä ja esineitä eri tavoin. Tämän puutteen vuoksi joissakin algoritmeissa on vaara, että ne toistavat ja jopa vahvistavat ihmisen ennakkoluuloja, erityisesti niitä, jotka vaikuttavat suojattuihin ryhmiin (Friedman & Nissenbaum, 1996; Lange & Duarte, 2017; Lee, Resnick & Barton, 2019).

Algoritmin vinouman vaikutukset: Algoritmin vinouma voi ilmetä monin eri tavoin, ja sen seuraukset kohderyhmälle vaihtelevat. Seuraavat esimerkit havainnollistavat sekä erilaisia syitä että vaikutuksia, jotka joko tahattomasti soveltavat ryhmiin erilaista kohtelua tai tuottavat niihin tarkoituksellisesti epäyhtenäisiä vaikutuksia (Lee, Resnick & Barton, 2019):

- *Vinoumat verkossa toimivissa rekrytointityökaluissa* - Verkkokauppayhtiö Amazon lopetti hiljattain rekrytointialgoritmin käytön havaittuaan, että se on sukupuolittunut. Tekoälyohjelmisto rankaisi kaikista ansioluetteloista, joiden tekstissä oli sana "naisten", ja alensi naisten yliopistoissa opiskelleiden naisten ansioluetteloita, mikä johti sukupuoleen perustuvaan vinoumaan.
- *Sanayhteyksien vinoumat* - Princetonin yliopiston tutkijat analysoivat ja yhdistivät 2,2 miljoonaa sanaa koneoppimisen tekoälyohjelmiston avulla. He havaitsivat, että eurooppalaiset nimet koettiin miellyttävämmiksi kuin afroamerikkalaisten nimet ja että sanat "nainen" ja "tyttö" yhdistettiin todennäköisemmin taiteisiin kuin tieteeseen ja matematiikkaan, jotka todennäköisimmin yhdistettiin miehiin.
- *Vinoumat verkkomainoksissa* - Harvardin tutkija ja Federal Trade Commissionin (FTC) entinen teknologiapäällikkö Latanya Sweeney havaitsi, että afroamerikkalaisille nimille tehtyt verkkohakukyselyt palauttavat todennäköisemmin pidätystietoja tarjoavan palvelun mainoksia kyseiselle henkilölle kuin valkoisille nimille tehtyt mainostulokset.
- *Kasvontunnistusteknologian vinoumat* - MIT:n tutkija Joy Buolamwini havaitsi, että kolmen kaupallisesti saatavilla olevan kasvontunnistusohjelmiston algoritmit eivät tunnistanee

tummaihoisia. Yleisesti ottaen useimmat kasvontunnistuksen harjoitusaineistot ovat arviolta yli 75-prosenttisesti miehiä ja yli 80-prosenttisesti valkoihoisia. Kun kuvassa esiintyvä henkilö oli valkoinen mies, ohjelmisto tunnisti henkilön 99 prosentissa tapauksista tarkasti mieheksi.

- *Rikosoikeuden algoritmeissa esiintyvä vinouma* - COMPAS-algoritmi (Correctional Offender Management Profiling for Alternative Sanctions), jota tuomarit käyttävät ennustamaan, pitäisikö syytetyt pidättää tai vapauttaa takuita vastaan oikeudenkäynnin ajaksi, osoittautui ProPublican raportin mukaan vääristyneeksi afroamerikkalaisia kohtaan.

5.2. Tietoisuus algoritmien vinouman kompleksisuudesta ja vivahteista

On tietoinen algoritmien vinoumien kompleksisuudesta ja vivahteista. Algoritmien vinouma, sen syiden ja vaikutusten analysointi. Analysoidaan erilaisia algoritmisen vääristelyn tyyppejä ja niiden perimmäisiä syitä. Arvioidaan algoritmisen vääristelyn vaikutuksia. Ymmärtää algoritmisen vinouman tunnistamiseen, mittaamiseen ja lieventämiseen liittyvät rajoitukset.

Algoritmien vinouman tyypit ja syyt: Algoritmin vinoumilla on eri muotoja. Friedmanin ja Nissenbaumin (1996) mukaan algoritmisen harhan eri tyypit ja syyt voidaan jakaa kolmeen luokkaan:

1. *Vinoumat* juontavat juurensa yhteiskunnallisiin instituutioihin, käytäntöihin ja asenteisiin. Kun tietokonejärjestelmissä ilmenee vinoumia, jotka ovat olemassa itsenäisesti ja yleensä ennen järjestelmän luomista, järjestelmä on esimerkki vinoumasta. Vinoumia voi esiintyä järjestelmässä joko yksilöiden tai instituutioiden nimenomaisen ja tietoisin toiminnan kautta tai epäsuorasti ja tiedostamatta, jopa parhaista aikomuksista huolimatta.

2. *Tekninen vinouma* johtuu teknisistä rajoituksista tai teknisistä näkökohdista. Teknisen vinouman lähteitä löytyy useista suunnitteluprosessin osa-alueista, kuten tietokonetyökalujen, kuten laitteistojen, ohjelmistojen ja oheislaitteiden rajoituksista, sosiaalisten merkitysten antamisesta kontekstista irrallseen kehitetyille algoritmeille, pseudosatunnaislukujen tuottamisen puutteista ja yrityksestä tehdä inhimillisistä konstruktioista tietokoneiden käyttökelpoisia, kun kvantifioidaan kvalitatiivista, diskretisoidaan jatkuvaa tai formalisoidaan epämuodollista.

3. *Kehittyvät vinoumat* syntyvät todellisten käyttäjien kanssa tapahtuvassa käytössä. Tämä ennakkoluulo syntyy tyypillisesti joskus suunnittelun

valmistumisen jälkeen yhteiskunnallisen tiedon, väestön tai kulttuuristen arvojen muuttumisen seurauksena. Käyttöliittymät ovat todennäköisesti erityisen alttiita ilmeneville vinoumille, sillä käyttöliittymät pyrkivät suunnittelussaan heijastamaan tulevien käyttäjien kykyjä, luonnetta ja tapoja. Näin ollen käyttöyhteyden muuttuminen voi hyvinkin aiheuttaa vaikeuksia uusille käyttäjille. Kehittyvät vinoumat johtuvat siitä, että yhteiskunnassa syntyy uutta tietoa, jota ei voida sisällyttää järjestelmään tai jota ei sisällytetä järjestelmään, tai kun järjestelmää käyttävä väestö eroaa joltain merkittävältä osin suunnittelussa käyttäjiksi oletetusta väestöstä, toisin sanoen käyttäjien ja järjestelmän suunnittelun välillä on epäsuhta.

5.3. Tunnistaa käytettävissä olevat strategiat ja välineet, joilla voidaan puuttua algoritmin vinoumaan liittyviin ongelmiin käytännössä

Tunnistetaan käytettävissä olevat strategiat ja välineet, joilla voidaan puuttua algoritmin vinoumaan liittyviin ongelmiin, mukaan lukien oikeudenmukaisuustietoinen koneoppiminen ja vinouman tarkastukset. Tarkastellaan ja yhdistetään käytettävissä olevia strategioita ja välineitä algoritmin vinoumaan liittyvien ongelmien käsittelemiseksi käytännössä. Kehitetään taitoja käyttää välineitä ja tekniikoita, joilla voidaan havaita, mitata ja lieventää algoritmista vääristymää tekoälyjärjestelmissä. Integroidaan tieteidenvälisiä lähestymistapoja, kuten alan asiantuntijoiden ja eri näkökulmien panos, tekoälyjärjestelmien oikeudenmukaisuuden parantamiseksi.

Strategiat ja välineet algoritmin vinoumaan liittyvien ongelmien ratkaisemiseksi: Algoritmin vinouman käsittely edellyttää huolellista harkintaa algoritmin kehittämisen eri vaiheissa. Siihen kuuluu monipuolisen ja edustavan harjoitteluaineiston varmistaminen, algoritmin suunnittelussa ja päätöksentekoprosessissa esiintyvien vinoumien tunnistaminen ja lieventäminen sekä testaus ja arviointi vinoumiin liittyvien ongelmien havaitsemiseksi ja korjaamiseksi. Syrjivien tulosten minimoimiseksi on tehty joitakin ponnisteluja kehittämällä kehyksiä ja suuntaviivoja algoritmien vinoumien lieventämiseksi, mukaan lukien algoritmijärjestelmien avoimuuden ja vastuullisuuden lisääminen, sääntelytoimenpiteet ja oikeudenmukaisuustietoisten koneoppimistekniikoiden käyttöönotto (Lee, Resnick & Barton, 2019).

4.2. CHARLIE Työpaketti 2 - Eettisen tekoälyn mikrokredentiaalikurssi EQF4

Tämä kurssi on suunniteltu antamaan kattava käsitys algoritmin vinoumista ja niiden vaikutuksista nykypäivän tietovetoiseen yhteiskuntaan. Koska algoritmeilla on merkittävä rooli elämämme muokkaamisessa ja päätöksenteossa eri aloilla, kuten terveydenhuollossa, koulutuksessa, rahoituksessa ja julkishallinnossa, ammattilaisten ja tutkijoiden/opiskelijoiden on tärkeää kehittää vankka käsitys näiden järjestelmien mahdollisista vääristymistä.

Tämän kurssin tavoitteena on antaa osallistujille tiedot ja taidot, joita tarvitaan algoritmin vinoumien tunnistamiseen, lieventämiseen ja poistamiseen. Teoreettisten perusteiden ja käytännön sovellusten yhdistelmän avulla kurssilaiset saavat käsityksen algoritmin vinoumiin liittyvistä eettisistä, sosiaalisista ja teknisistä näkökohdista.

Nämä tavoitteet edistävät yleisiä tavoitteita asettamalla kulmakivet seuraaville toimille, jotta opettajilla, ohjaajilla ja oppijoilla olisi selkeä näkemys eri oppimispolkujen tavoitteista, ja jotta korkea-asteen koulutuksen, aikuiskoulutuksen ja nuorten opetuksen valmiuksia tarjota yhteiskunnan tarpeita vastaavia, mutta myös oppijoiden oppimistarpeiden mukaan räätälöityjä oppimismahdollisuuksia voitaisiin parantaa. Eettisen tekoälyn mikrokredentiaalikurssin EQF4 odotetut tuotokset ovat seuraavat:

- 1- Kehitetään perusteellinen ymmärrys algoritmin vinoumasta, sen syistä ja seurauksista yksilöille ja yhteiskunnalle.
 - 1.1. Hankitaan tietoa algoritmin vinouman määritelmästä, lähteistä ja ilmenemismuodoista.
 - 1.2. Ymmärtää vinoutuneiden algoritmien yhteiskunnalliset ja yksilölliset vaikutukset.
2. Edistetään tietoisuutta siitä, että tekoälyä kehitettäessä on noudatettava eettistä periaatetta, jonka mukaan tekoälyä ei saa vahingoittaa.
 - 2.1. Tutustutaan vinoutuneisiin algoritmeihin liittyviin riskeihin ja haittoihin.
 - 2.2. Kehitetään strategioita haittojen minimoimiseksi ja tekoälyn eettisen kehityksen edistämiseksi.
3. Ymmärtää vastuullisuuden merkitys tekoälyjärjestelmissä ja tutkia asiaa koskevia oikeudellisia ja eettisiä puitteita.
 - 3.1. Tutkitaan eri sidosryhmien roolia tekoälyn vastuullisuudessa.
 - 3.2. Opitaan parhaat käytännöt vastuullisuuden varmistamiseksi tekoälyn kehittämisessä.

4. Saada tietoa tekoälyjärjestelmien avoimuuden käsitteestä ja sen merkityksestä algoritmisessa päätöksenteossa.
 - 4.1. Tutkitaan menetelmiä ja välineitä tekoälyn avoimuuden lisäämiseksi.
 - 4.2. Ymmärtää monimutkaisten algoritmien ymmärrettävyyden parantamiseen liittyvät haasteet ja rajoitukset.
5. Tutkitaan tekoälyn, ihmisoikeuksien ja oikeudenmukaisuuden yhtymäkohtia ja niiden vaikutusta tekoälyn eettiseen kehittämiseen.
 - 5.1. Tutkitaan, miten vinoutuneet algoritmit voivat vaikuttaa ihmisoikeuksiin, kuten syrjimättömyyteen, yksityisyyteen ja sananvapauteen.
 - 5.2. Opitaan strategioita, joilla varmistetaan oikeudenmukaisuus ja tasapuolisuus tekoälyn kehittämisessä ja käyttöönotossa.
6. Soveltaa eettisiä periaatteita tekoälyn kehittämisessä ja käyttöönotossa käytännönläheisten lähestymistapojen ja todellisten skenaarioiden avulla.
 - 6.1. Tutustutaan erilaisiin eettisiin kehyksiin ja ohjeisiin ja niiden soveltamiseen tekoälyjärjestelmiin.
 - 6.2. Ymmärtää sidosryhmien sitouttamisen, tieteidenvälisen yhteistyön ja eettisesti kestävien tekoälyn kehitysprosessien toteuttamisen merkityksen.

Kurssin jälkeen osallistujat ymmärtävät kattavasti algoritmin vinoumia, niiden mahdollisia vaikutuksia eri aloihin sekä käytettävissä olevia strategioita ja välineitä, joilla näihin vinoumiin voidaan puuttua. Tämä tieto on korvaamatonta ammattilaisille ja korkeakouluopiskelijoille/opiskelijoille, jotka työskentelevät aloilla, joilla algoritmeja hyödynnetään päätöksenteossa, ja se auttaa varmistamaan tasa-arvoisemmat ja oikeudenmukaisemmat tulokset tietoon perustuvassa maailmassa.

4.2.1. Eettisen tekoälyn mikrokredentiaalikurssin EQF4 sisältö

Kurssi rakentuu kuudesta osaamisyksiköstä (CU), joista jokainen on suunniteltu antamaan osallistujille tiedot ja taidot, joita tarvitaan algoritmin vinouman haasteiden ja mahdollisuuksien hallintaan.

CU1 - Mitä on algoritmin vinouma? Oppijat tutustuvat algoritmin vinouman käsitteeseen ja sen eri ilmenemismuotoihin. Käsiteltäviä aiheita ovat muun muassa algoritmin vinouman määritelmä, algoritmin vinouman syyt ja lähteet sekä vinoutuneista algoritmeista yksilöille ja yhteiskunnalle mahdollisesti aiheutuvat seuraukset.

CU2 – Vahingon välttäminen (Non-maleficence) Tässä yksikössä keskitytään non-maleficence-periaatteeseen, jossa korostetaan vahingon välttämisen tärkeyttä tekoälyjärjestelmien kehittämisessä ja

käyttöön otossa. Oppijat tutustuvat vinoutuneisiin algoritmeihin liittyviin mahdollisiin riskeihin ja haittoihin sekä strategioihin, joilla haitat voidaan minimoida ja varmistaa tekoälyn eettinen kehittäminen.

CU3 - Vastuullisuus Vastuullisuus-yksikössä oppijat tarkastelevat, miten tärkeää on luoda selkeät vastuu- ja tilivelvollisuuslinjat tekoälyjärjestelmien kehittämisessä ja käytössä. Käsiteltäviä aiheita ovat muun muassa oikeudelliset ja eettiset vastuukehykset, eri sidosryhmien rooli ja parhaat käytännöt vastuullisuuden varmistamiseksi tekoälyn kehittämisessä.

CU4 - Läpinäkyvyys Tässä yksikössä syvennyttään tekoälyjärjestelmien läpinäkyvyyden käsitteeseen ja tarkastellaan avoimuuden, viestinnän ja selitettävyyden merkitystä algoritmisessa päätöksenteossa. Opiskelijat tutustuvat menetelmiin ja työkaluihin, joilla voidaan lisätä tekoälyn läpinäkyvyyttä, sekä haasteisiin ja rajoituksiin, jotka liittyvät monimutkaisten algoritmien tekemiseen ymmärrettävämmiksi.

CU5 - Ihmisoikeudet ja oikeudenmukaisuus Ihmisoikeudet ja oikeudenmukaisuus -yksikössä opiskelijat tutkivat tekoälyn, ihmisoikeuksien ja oikeudenmukaisuuden yhtymäkohtia. He tutkivat, miten puolueelliset algoritmit voivat vaikuttaa ihmisoikeuksiin, kuten syrjimättömyyteen, yksityisyyteen ja sananvapauteen. Opiskelijat tutustuvat myös strategioihin, joilla varmistetaan oikeudenmukaisuus ja tasapuolisuus tekoälyn kehittämisessä ja käyttöön otossa.

CU6 - Tekoälyn etiikka, käytännönläheinen lähestymistapa Tässä yksikössä keskitytään eettisten periaatteiden käytännön soveltamiseen tekoälyn kehittämisessä ja käyttöön otossa. Opiskelijat tutustuvat erilaisiin eettisiin kehyksiin ja ohjeisiin ja oppivat soveltamaan niitä tekoälyjärjestelmiin liittyviin todellisiin skenaarioihin. Yksikössä käsitellään myös sidosryhmien sitouttamisen, monitieteisen yhteistyön ja eettisten tekoälyn kehitysprosessien toteuttamisen merkitystä.

4.2.2. Kohderyhmän kuvaus, EQF4-ominaisuudet ja piirteet

Eurooppalainen tutkintojen viitekehys (EQF) on yhteinen viitekehys, joka yhdistää eri Euroopan maiden tutkintojärjestelmät ja helpottaa yksilöiden tutkintojen tunnustamista kaikkialla Euroopassa (Euroopan komissio, 2023). EQF koostuu kahdeksasta viitetasosta, joista kukin on määritelty kuvaajilla, jotka osoittavat kyseisellä tasolla odotetut oppimistulokset, taidot, pätevyydet ja itsenäisyyden. EQF-taso 4 (EQF4) vastaa tutkintoja, jotka on suoritettu keskiasteen jälkeisen muun kuin korkea-asteen koulutuksen tai ammatillisen

tutkinnon tasolla Euroopan korkeakoulutusalueella (EHEA). EQF4:n ominaispiirteitä ja piirteitä ovat (Euroopan komissio, 2023; CEDEFOP, 2023):

- Faktatieto ja teoreettinen tieto: EQF4-tutkinnoissa edellytetään, että oppijoilla on laaja valikoima tosiasioihin perustuvaa ja teoreettista tietoa opiskelualalta tai ammatista. Tällainen tietämys on välttämätöntä, jotta oppija voi ymmärtää keskeisiä periaatteita, käsitteitä ja prosesseja, jotka ovat hänen ammattialansa kannalta olennaisia.
- Kognitiiviset taidot: EQF4:ssä oppijoiden tulisi kehittää kykyä soveltaa tietojaan käytännön tilanteissa, analysoida tilanteita ja ratkaista ongelmia perusmenetelmiä ja -välineitä käyttäen. Tähän sisältyy kriittinen ajattelu, ongelmanratkaisu ja päätöksentekotaidot perustavammalla tasolla.
- Käytännön taidot: EQF4-tutkintoihin kuuluu erilaisten käytännön taitojen kehittäminen, jonka avulla oppijat voivat suorittaa tehtäviä ja toimia valitsemallaan alalla. Näihin taitoihin voi kuulua teknisiä taitoja, asiaankuuluvien työkalujen ja laitteiden käyttöä tai kykyä toteuttaa tiettyjä prosesseja ja menettelyjä.
- Itsenäisyys ja vastuu: EQF4:n oppijoiden odotetaan osoittavan työssään kohtalaista itsenäisyyttä ja vastuullisuutta. Tähän kuuluu vastuun ottaminen omasta oppimisestaan ja kehittymisestään, työskentely valvonnan tai ohjauksen alaisena ja päätösten tekeminen vakiintuneiden ohjeiden ja menettelyjen perusteella.
- Viestintä ja yhteistyö: EQF4-tutkinnot edellyttävät, että oppijat kehittävät tehokkaita viestintä- ja yhteistyötaitoja, joiden avulla he voivat vaihtaa tietoja, väitteitä tai ehdotuksia muiden kanssa. Tähän kuuluvat kirjalliset, suulliset ja ihmissuhdetaidot sekä kyky työskennellä tehokkaasti tiimeissä tai ryhmissä.
- Elinikäinen oppiminen: EQF4:ssä oppijoiden tulisi kehittää kykyä osallistua elinikäiseen oppimiseen, pohtia omia oppimiskokemuksiaan ja yksilöidä alueita, joilla he voivat kehittyä edelleen. Tähän sisältyy kyky sopeutua uusiin tilanteisiin, teknologioihin tai ammatillisiin ympäristöihin ja osallistua jatkuvaan ammatilliseen kehitykseen.

Kurssin aikana kehitettävät taidot ovat EQF4-kuvaajien mukaisia, ja niissä keskitytään siihen, että oppijat saavat monenlaisia tietoja, kognitiivisia ja käytännön taitoja, itsenäisyyttä, vastuullisuutta, viestintää, yhteistyötä ja

sitoutumista elinikäiseen oppimiseen. Algoritmista ennakkoluulottomuutta käsittelevän EQF4-kurssin osaamisalueisiin kuuluvat:

- Faktatieto ja teoreettinen tietämys: 1.1. Ymmärtää algoritmien, datan ja tekoälyn peruskäsitteet. 1.2. Tunnistaa algoritmien roolin eri aloilla ja vääristyneen päätöksenteon mahdolliset seuraukset.
- Kognitiiviset taidot: 2.1. Tunnistaa ja analysoida yksinkertaisia esimerkkejä algoritmista vinoumasta ja sen vaikutuksista reaali maailman skenaarioissa. 2.2. Soveltaa kriittistä ajattelua arvioidakseen tekoälyjärjestelmien oikeudenmukaisuutta ja eettisiä näkökohtia tietyssä kontekstissa.
- Käytännön taidot: 3.1. käyttää perustyökaluja ja -menetelmiä mahdollisten vinoumien tunnistamiseksi tietokokonaisuuksissa ja algoritmisissa prosesseissa. 3.2. Tutkitaan yksinkertaisia tekniikoita tekoälyjärjestelmissä esiintyvien vinoumien käsittelemiseksi ja lieventämiseksi.
- Itsenäisyys ja vastuu: 4.1. Ottaa vastuun omasta oppimisestaan ja kehittymisestään algoritmiseen ennakkoluuloon liittyvissä asioissa. 4.2. Pohtia algoritmisen päätöksenteon eettisiä vaikutuksia omassa työssään tai opiskelualallaan.
- Viestintä ja yhteistyö: 5.1. Tehokas viestintä ja keskustelu algoritmiseen vinoumaan liittyvistä kysymyksistä vertaisten, opettajien ja muiden sidosryhmien kanssa. 5.2. tehdä yhteistyötä muiden kanssa algoritmiseen vinoumaan liittyvien tapausten tutkimiseksi ja käsittelemiseksi ryhmähankkeissa tai tapaustutkimuksissa.
- Elinikäinen oppiminen: 6.1. Omien oppimiskokemusten pohtiminen ja algoritmisen ennakkoluulon alalla jatkokehittämisen kohteiden määrittäminen. 6.2. Osoitetaan sitoutuminen pysymään ajan tasalla tekoälyn etiikan ja algoritmisen vinouman alalla tapahtuneista edistysaskelista ja esiin nousevista kysymyksistä.

Keskittymällä näihin kompetensseihin algoritmiseen vinoumaan perustuva EQF4-mikrokredentiaalikurssi antaisi oppijoille vankan perustan alalle ja antaisi heille tiedot, taidot ja ymmärryksen, joita tarvitaan, jotta he voivat selviytyä vinoutuneisiin algoritmeihin ja tekoälyjärjestelmiin liittyvistä eettisistä haasteista tulevassa urassaan tai korkeakouluopinnoissaan.

4.2.3. Yleinen kompetenssimatriisi EQF4:n mikrokredentiaalikurssia varten

OPPIMISTULOKSET		
Kurssin suoritettuaan oppija osaa seuraavat asiat		
TIETOJA	TAIDOT	ASENTEET
1. Ymmärtää algoritmien, tekoälyn ja koneoppimisen peruskäsitteet ja niiden sovellukset eri aloilla. 2. Määrittää algoritmien vinouma ja tunnistaa sen mahdolliset lähteet, mukaan lukien data, mallien suunnittelu ja toteutus. 3. Tunnistaa algoritmien vinouman seuraukset yksilöille, yhteisöille ja koko yhteiskunnalle. 4. Selittää tekoälyjärjestelmien käyttöön liittyvät eettiset näkökohdat, mukaan lukien oikeudenmukaisuus, läpinäkyvyys ja vastuullisuus. 5. Tunnistaa reaali maailman esimerkkejä ja tapauksia, jotka havainnollistavat algoritmien vinouman vaikutusta ja sen eettisiä seurauksia. 6. Arvostaa monipuolisten ja edustavien tietokokonaisuuksien merkitystä tekoälyjärjestelmien kehittämisessä. 7. Tutkia erilaisia strategioita ja tekniikoita tekoälyjärjestelmissä esiintyvien algoritmien vinoumien lieventämiseksi ja käsittelemiseksi. 8. Keskustella lainsäädännön, sääntelyn ja alan standardien roolista algoritmien vinouman käsittelyssä ja tekoälyn eettisen kehityksen edistämisessä. 9. Tutkia haasteita ja mahdollisuuksia, jotka liittyvät tekoälyjärjestelmien hyötyjen ja algoritmiseen vinoumaan liittyvien mahdollisten riskien tasapainottamiseen. 10. Tunnistaa tieteiden välisen yhteistyön merkityksen oikeudenmukaisen ja eettisten	1. Analysoida ja arvioida algoritmeja mahdollisten vinoumien ja eettisten vaikutusten kannalta. 2. Soveltaa kriittisen ajattelun taitoja arvioidakseen tekoälyjärjestelmien oikeudenmukaisuutta ja tasapuolisuutta reaali maailman tilanteissa. 3. Käyttää tilastollisia perustekniikoita havaitakseen ja mitataksaan vinoumia tietoaaineistoissa. 4. Soveltaa strategioita, joilla voidaan käsitellä ja lieventää vinoumia tekoälyjärjestelmien kehittämisessä ja käyttöönotossa. 5. Viestiä tehokkaasti algoritmien vinoumista, niiden seurauksista ja mahdollisista ratkaisuista sekä teknisille että muille kuin teknisille kohderyhmille. 6. Tehokas yhteistyö erilaisissa tiimeissä tekoälyjärjestelmien kehittämiseksi ja arvioimiseksi ottaen huomioon eettiset ja oikeudenmukaisuusnäkökohdat. 7. Soveltaa asiaankuuluvaa lainsäädäntöä, määräyksiä ja alan standardeja tekoälyjärjestelmien kehittämiseen ja käyttöönottoon. 8. Mukauttaa ja soveltaa parhaita käytäntöjä eettiseen tekoälyn kehittämiseen eri aloilla ja toimialoilla. 9. Osoittaa ongelmanratkaisutaitoja käsitellessään reaali maailman tapauksia, joissa esiintyy algoritmista vinoumaa, ja kehittää mahdollisia ratkaisuja. 10. Päivittää jatkuvasti tietoaan ja taitojaan algoritmien	1. Arvostaa eettisten näkökohtien, oikeudenmukaisuuden ja tasapuolisuuden merkitystä tekoälyjärjestelmien kehittämisessä ja käyttöönotossa. 2. Osoittaa sitoutumista elinikäiseen oppimiseen ja jatkuvaan parantamiseen tekoälyn eettisten periaatteiden ymmärtämisessä ja soveltamisessa. 3. Omaksuu vastuullisen ajattelutavan tekoälyjärjestelmien kehittämisessä, käytössä ja arvioinnissa. 4. Suhtautuu ennakkoivasti tekoälysovellusten mahdollisten vinoumien ja eettisten ongelmien tunnistamiseen ja käsittelyyn. 5. Tunnustaa tieteiden välisen yhteistyön ja avoimen viestinnän merkityksen algoritmiseen vinoumaan ja tekoälyn eettiseen kehittämiseen puuttumisessa. 6. Arvostaa avoimuuden, vastuuvollisuuden ja sidosryhmien osallistumisen merkitystä tekoälyjärjestelmien kehittämisessä. 7. Kunnioittaa tekoälyjärjestelmien vaikutuspiiriin kuuluvien henkilöiden erilaisia näkökulmia ja kokemuksia ja ottaa huomioon heidän panoksensa oikeudenmukaisen ja puolueettomien ratkaisujen suunnittelussa ja täytäntöönpanossa. 8. Edistää tekoälyn kehittämisessä osallistavaa lähestymistapaa, jossa otetaan huomioon eri sidosryhmien, myös syrjäytyneiden ja aliedustettujen yhteisöjen, tarpeet ja huolenaiheet. 9. Tunnustaa tekoälyjärjestelmien rajoitukset ja jatkuvan arvioinnin ja parantamisen tärkeyden mahdollisten haittojen minimoimiseksi. 10. Edistää tekoälyn eettistä kehittämistä ja tietoisuutta algoritmista vinoumasta ja sen seurauksista ammatillisissa ja yhteiskunnallisissa yhteyksissä.

tekoälyjärjestelmien kehittämisessä ja toteuttamisessa, mukaan lukien eri taustoista ja eri aloilta tulevien sidosryhmien osallistuminen.	vinouman ja eettisen tekoälyn kehittämisen alalla pysyäkseen ajan tasalla kehittyvistä suuntauksista ja teknologioista.	
---	---	--

4.2.4. EQF4-mikrokredentiaalikurssin kunkin osaamisyksikön osaamismatriisi

CU1 - Mitä on algoritmisen vinouma? EQF4		
OPPIMISTULOKSET		
Kurssin suoritettuaan oppija osaa seuraavat asiat		
TIEDOT	TAIDOT	ASENTEET
Teoreettinen/tosiasiallinen tietämys seuraavista: CUK1.1. Määrittellä algoritmisen vinouma ja ymmärtää sen peruskäsitteet, mukaan lukien tekijät, jotka vaikuttavat tekoälyjärjestelmien vinoutuneisiin tuloksiin. CUK1.2. Tunnistaa erityyppiset algoritmiset vinoumat, kuten dataan perustuva vinouma, malliin perustuva vinouma ja ihmislähtöinen vinouma, ja tunnistaa, miten ne voivat ilmetä tekoälysovelluksissa. CUK1.3. Ymmärtää algoritmisen vinouman todelliset vaikutukset eri aloilla, mukaan lukien mahdolliset sosiaaliset, eettiset ja oikeudelliset seuraukset, jotka liittyvät tekoälyjärjestelmien vinoumiin.	CUS1.1. Analysoida tekoälyjärjestelmiä ja -sovelluksia mahdollisten vinoumien havaitsemiseksi käyttäen algoritmisen vinouman peruskäsitteiden tuntemustaan. CUS1.2. Luokitella erityyppisiä algoritmisia vinoumia reaali maailman esimerkeissä käyttäen tietämystään dataan perustuvista, malliin perustuvista ja ihmislähtöisistä vinoumista. CUS1.3. Arvioida algoritmisen vinouman vaikutusta eri aloihin ja sidosryhmiin ottaen huomioon mahdolliset sosiaaliset, eettiset ja oikeudelliset seuraukset, jotka liittyvät tekoälyjärjestelmien vinoumiin.	CUA1.1. Osoittaa lisääntyntä tietoisuutta algoritmisen vinouman mahdollisista seurauksista ja omaksuu kriittisen ajattelutavan tekoälyjärjestelmiä ja -sovelluksia käyttäessään. CUA1.2. Hyväksyy oikeudenmukaisuuden ja tasapuolisuuden merkityksen tekoälyjärjestelmissä ja tunnustaa, että on tärkeää käyttää erilaisia näkökulmia ja osallistavia suunnitteluprosesseja ennakkoluulojen minimoimiseksi. CUA1.3. Sitoutuu jatkuvaan oppimiseen ja pysyy ajan tasalla tekoälyn etiikan, algoritmisen vinouman ja vastuullisten tekoälykäytäntöjen viimeisimmästä kehityksestä ja pyrkii edistämään myönteisesti puolueettomien tekoälyjärjestelmien kehittämistä ja käyttöä.

CI OSAAMISTULOSTEN KUVAUS

Algoritmista vinoumaa käsittelevällä EQF4-tason kurssilla opiskelijat hankkivat perustiedot alalta keskittyen peruskäsitteiden ymmärtämiseen, erityyppisten vinoumien tunnistamiseen ja algoritmisen vinouman todellisten vaikutusten tunnistamiseen eri aloilla. Tämän kurssin osaamistavoitteisiin kuuluvat:

CU1.1. Algoritmisen vinouman määrittely: Opiskelijat oppivat määrittelemään algoritmisen vinouman ja ymmärtämään sen peruskäsitteet. He tutkivat tekijöitä, jotka edistävät tekoälyjärjestelmien vinoutuneita tuloksia, kuten vääristyneitä tiedonkeruumenetelmiä, vinoutuneita harjoitustietoja ja ihmisen päätöksentekoa. Tämä perustieto auttaa opiskelijoita tunnistamaan, miten algoritmia

vinoumaa voi esiintyä ja miten se voi vaikuttaa tekoälysovelluksiin, esimerkiksi kasvojentunnistusjärjestelmiin, jotka tunnistavat virheellisesti tiettyjä väestöryhmiä.

CU1.2. Algoritmien vinoumien tunnistaminen: Opiskelijat tutustuvat erilaisiin algoritmisiin vinoumiin, kuten dataan perustuviin vinoumiin, malliin perustuviin vinoumiin ja ihmislähtöisiin vinoumiin. Esimerkkien ja tapaustutkimusten avulla he oppivat, miten nämä vinoumat voivat ilmetä tekoälysovelluksissa, ja ymmärtävät, miten tärkeää on puuttua niihin, jotta voidaan varmistaa oikeudenmukaiset ja puolueettomat tekoälyjärjestelmät. Opiskelijat tarkastelevat esimerkiksi sitä, miten dataan perustuva vinouma voi johtua epäedustavasta harjoitusdatasta, mikä johtaa vinoutuneisiin ennusteisiin esimerkiksi luottopisteytyksessä tai työnhakijoiden seulonnassa.

CU1.3. Algoritmisen vinouman todelliset vaikutukset: Opiskelijat tutkivat algoritmisen vinouman todellisia vaikutuksia eri aloilla, kuten terveydenhuollossa, rahoituksessa ja rikosoikeudessa. Opiskelijat tarkastelevat tekoälyjärjestelmiin liittyviä mahdollisia sosiaalisia, eettisiä ja oikeudellisia seurauksia ja oppivat ymmärtämään, että algoritmisen vinouman minimoiminen on tarpeen kielteisten tulosten estämiseksi ja tekoälysovellusten oikeudenmukaisuuden ja tasapuolisuuden edistämiseksi. Esimerkkeinä voidaan mainita, miten tekoälyjärjestelmien vinoumat terveydenhuollossa voi johtaa väärin diagnooseihin tai riittämättömiin hoitoihin tietyille väestöryhmille tai miten algoritmisen vinouman rikosoikeudellisissa järjestelmissä voi johtaa epäoikeudenmukaiseen tuomitsemiseen tai yksilöiden profilointiin.

CU2 - Vahingon välttäminen (Non-maleficence) EQF4

OPPIMISTULOKSET

Kurssin suoritettuaan oppija osaa seuraavat asiat		
TIEDOT	TAIDOT	ASENTEET
Teoreettinen/tosiasiallinen tietämys seuraavista: CUK2.1. Esittelee non-maleficence-periaate: Esittelee oppilaille non-maleficence-periaatteen peruskäsitteitä ja painota, miten tärkeää on välttää haittoja tekoälyjärjestelmiä luotaessa ja käytettäessä ja miten tämä ajatus edistää vastuullista tekoälyn kehittämistä. CUK2.2. Tutki tekoälyn vinoutumien aiheuttamia mahdollisia haittoja: Ohjaa oppilaita tunnistamaan eri tapoja, joilla vinoutuneet tekoälyjärjestelmät voivat aiheuttaa haittaa, kuten syrjintää tai yksityisyyden loukkaamista, ja havainnollistamaan todellisten esimerkkien avulla, miten tärkeää on puuttua algoritmien vinoutumiin. CUK2.3. Tutki strategioita tekoälyjärjestelmien muuttamiseksi vähemmän haitallisiksi: Kerro oppilaille yksinkertaisista strategioista, joilla tekoälyjärjestelmistä voidaan tehdä vähemmän haitallisia, kuten oikeudenmukaisuuden, vastuullisuuden ja avoimuuden edistäminen tekoälyn kehittämisessä ja yhteistyön edistäminen eri alojen asiantuntijoiden kanssa.	CUS2.1. Selittää non-maleficence-käsite tekoälyn kehittämisen yhteydessä. CUS2.2. Tunnistaa tekoälyjärjestelmien mahdolliset haitat ja tunnustaa non-maleficence-periaatteen merkityksen tekoälyn vastuullisessa kehittämisessä. CUS2.3. Analysoida erityyppisiä haittoja, jotka voivat aiheutua tekoälyjärjestelmien puolueellisuudesta. CUS 2.4. Arvioida todellisia esimerkkejä tekoälyjärjestelmien vinoutumista ja niiden seurauksia. CUS2.5. Tunnistaa ja soveltaa strategioita tekoälyjärjestelmien haittojen vähentämiseksi, kuten oikeudenmukaisuuden, vastuullisuuden ja avoimuuden edistäminen. CUS2.6. Tehokas yhteistyö eri alojen asiantuntijoiden kanssa vinoutuneiden tekoälyjärjestelmien aiheuttamien haittojen käsittelemiseksi ja lieventämiseksi.	CUA2.1. Kasvattaa vastuuntuntoa siitä, että tekoälyn kehittämisessä sovelletaan non-maleficence-periaatetta. CUA2.2. Arvostaa eettisten näkökohtien merkitystä tekoälyjärjestelmissä, erityisesti mahdollisten haittojen välttämiseksi. CUA2.3. Kehittää empatiaa yksilöitä ja yhteisöjä kohtaan, joihin tekoälyjärjestelmien aiheuttamat haitat vaikuttavat. CUA2.4. Edistää sitoutumista siihen, että tekoälyn kehityksessä käsitellään ja lievennetään algoritmista vinoumaa mahdollisten haittojen vähentämiseksi. CUA2.5. Omaksuu ennakoivan lähestymistavan sellaisten strategioiden toteuttamiseen, joilla vähennetään tekoälyjärjestelmien aiheuttamia haittoja. CUA2.6. Arvostaa monitieteisen yhteistyön ja erilaisten näkökulmien merkitystä käsiteltäessä ja lievennettäessä tekoälyjärjestelmien aiheuttamia haittoja.

OSAAMISTULOSTEN KUVAUS

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 todellisten vaikutusten tunnistamiseen.

Kurssin osaamistavoitteet ovat seuraavat:

CU2.1. Non-maleficence-periaate: Oppijat ymmärtävät non-maleficence-periaatteen peruskäsitteen ja korostavat, miten tärkeää

on välttää haittoja tekoälyjärjestelmiä luotaessa ja käytettäessä, ja miten tämä ajatus edistää vastuullista tekoälyn kehittämistä.

CU2.2. Vinoutuneesta tekoälystä mahdollisesti aiheutuvat haitat:

Opiskelijat tunnistavat eri tapoja, joilla vinoutuneet tekoälyjärjestelmät voivat aiheuttaa haittaa, kuten syrjintää tai yksityisyyden loukkaamista, ja käyttävät todellisia esimerkkejä havainnollistamaan algoritmiseen vinoumaan puuttumisen merkitystä.

CU2.3. Strategiat tekoälyjärjestelmien muuttamiseksi vähemmän haitallisiksi: Oppijat tutustuvat yksinkertaisiin strategioihin, joilla tekoälyjärjestelmistä voidaan tehdä vähemmän haitallisia, kuten oikeudenmukaisuuden, vastuullisuuden ja avoimuuden edistämiseen tekoälyn kehittämisessä ja yhteistyön edistämiseen eri alojen asiantuntijoiden kanssa.

CU3 - Vastuullisuus EQF4		
OPPIMISTULOKSET		
	Kurssin suoritettuaan oppija osaa seuraavat asiat	
TIEDOT	TAIDOT	ASENTEET
Teoreettinen/tosiasiallinen tietämys seuraavista: CUK3.1. Tekoälyn vastuullisuuden peruskäsite: Oppijat kuvaavat tekoälyn vastuullisuuden peruskäsitteen ja korostavat tekoälyn kehittäjien ja käyttäjien vastuuta varmistaa, että tekoälyjärjestelmät toimivat eettisesti ja mahdollisimman vähän haittaa aiheuttaen. CUK3.2. Vastuullisuuden rooli algoritmisen vinouman käsittelyssä: Oppijat tunnistavat, miten vastuullisuudella on ratkaiseva merkitys algoritmisen vinouman vaikutusten ehkäisemisessä ja lievittämisessä, ja miten vastuun ottaminen tekoälyjärjestelmistä voi johtaa parempaan päätöksentekoon ja oikeudenmukaisempiin tuloksiin. CUK3.3. Mekanismit tekoälyjärjestelmien vastuullisuuden varmistamiseksi: Opiskelijat tutustuvat yksinkertaisiin mekanismeihin, jotka voivat auttaa varmistamaan vastuullisuuden tekoälyjärjestelmissä, kuten ohjeisiin, verkkoetikettiin, hyväksyttävän käytön periaatteisiin (AUP) ja tekoälyn kehittäjille ja käyttäjille tarkoitettuihin määräyksiin.	CUS3.1. Selittää vastuun käsite tekoälyn kehittämisen ja käytön yhteydessä. CUS3.2. Tunnistaa tekoälyn kehittäjien ja käyttäjien vastuun eettisten ja mahdollisimman vähän haittaa aiheuttavien tekoälyjärjestelmien varmistamisessa. CUS3.3. Analysoida vastuullisuuden ja algoritmisen vinouman välistä suhdetta ja ymmärtää sen merkityksen tekoälyn kehittämisessä. CUS3.4. Hyödyntää vastuullisuuden käsitettä reaali maailman skenaarioissa päätöksenteon parantamiseksi ja oikeudenmukaisempien tulosten saavuttamiseksi. CUS3.5. Tunnistaa ja arvioida mekanismeja, joilla varmistetaan tekoälyjärjestelmien vastuullisuus, kuten ohjeet, verkkoetiikka, hyväksyttävän käytön periaatteet (AUP) ja tekoälyn kehittäjille ja käyttäjille annetut määräykset. CUS3.6. Valita ja mukauttaa näitä mekanismeja tekoälyn kehittämisessä ja käyttää niitä vastuullisuuden ja eettisen käyttäytymisen edistämiseksi.	CUA3.1. Kasvattaa vastuullisuutta tekoälyn eettisestä kehittämisestä ja käytöstä ja tiedostaa vastuullisuuden merkitys. CUA3.2. Arvostaa vastuullisuuden merkitystä haittojen minimoimisessa ja eettisen käyttäytymisen edistämässä tekoälyjärjestelmissä. CUA3.3. Arvostaa vastuuvastuullisuuden merkitystä algoritmiseen vinoumaan ja sen mahdollisiin seurauksiin puuttumisessa. CUA3.4. Edistää sitoutumista vastuun kantamiseen tekoälyjärjestelmistä paremman päätöksenteon ja oikeudenmukaisempien tulosten saavuttamiseksi. CUA3.5. Hyväksyy sellaisten mekanismien käyttöönoton ja toteuttamisen, joilla edistetään tekoälyjärjestelmien vastuullisuutta. CUA3.6. Arvostaa ohjeiden, verkkoetiikan, hyväksyttävän käytön periaatteiden ja säännösten merkitystä vastuullisuuden ja eettisen käyttäytymisen kulttuurin edistämässä tekoälyn kehittämisessä ja käytössä.

OSAAMISTULOSTEN KUVAUS

Kurssin osaamistavoitteet ovat seuraavat:

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 todellisten vaikutusten tunnistamiseen.

Kurssin osaamistavoitteet ovat seuraavat:

CU3.1. Tekoälyn vastuullisuuden peruskäsite: Oppijat oppivat tekoälyn vastuullisuuden peruskäsitteen. Tekoälyn vastuullisuus liittyy siihen oletukseen, että suunnittelijat, kehittäjät ja käyttöönottajat noudattavat standardeja ja lainsäädäntöä, joilla varmistetaan tekoälyn asianmukainen toiminta sen elinkaaren aikana (Fjeld et al. 2020). Oppijat tunnistavat tekoälyn kehittäjien ja käyttäjien vastuun eettisten ja mahdollisimman vähän haittaa aiheuttavien tekoälyjärjestelmien varmistamisessa. He kasvattavat vastuuntuntoa tekoälyn eettisestä kehittämisestä ja käytöstä, tunnistavat vastuullisuuden merkityksen ja arvostavat sen merkitystä haittojen minimoimisessa ja eettisen käyttäytymisen edistämisessä tekoälyjärjestelmissä.

CU3.2. Vastuullisuuden rooli algoritmisen vinouman torjunnassa: Oppijat tunnistavat syyt siihen, miksi vastuullisuus on ratkaisevassa asemassa algoritmisen vinouman vaikutusten ehkäisemisessä ja lieventämisessä. He oppivat analysoimaan vastuullisuuden ja algoritmisen vinouman välistä suhdetta ja arvostavat sitä, miten tärkeää on ottaa vastuu tekoälyjärjestelmistä paremman päätöksenteon ja oikeudenmukaisempien tulosten saavuttamiseksi.

CU3.3. Mekanismit tekoälyjärjestelmien vastuullisuuden varmistamiseksi. Opiskelijat tutustuvat yksinkertaisiin mekanismeihin, joiden avulla voidaan varmistaa vastuullisuus tekoälyjärjestelmissä. Opiskelijat ymmärtävät, että on tärkeää viestiä tehokkaasti sidosryhmien kanssa (esimerkiksi antamalla tietoa tekoälyjärjestelmän perusteluista ja logiikasta, kuten panoksista, tuotoksista ja prosesseista, joita käytetään päätösten tai suositusten tekemiseen). Esimerkki mekanismeista: ohjeet, verkkoetiikka, hyväksyttävän käytön periaatteet (AUP) ja määräykset tekoälyn kehittäjille ja käyttäjille.

CU4 - Läpinäkyvyys EQF4		
OPPIMISTULOKSET		
	Kurssin suoritettuaan oppija osaa seuraavat asiat	
TIEDOT	TAIDOT	ASENTEET
Teoreettinen/tosiasiallinen tietämys seuraavista: CUK4.1. Avoimuuden merkitys tekoälyjärjestelmissä: Opiskelijat oppivat tekoälyn läpinäkyvyyden peruskäsitteen ja sen merkityksen sen varmistamisessa, että tekoälyjärjestelmät ovat ymmärrettäviä, selitettävissä ja sidosryhmien saatavilla. CUK4.2. Läpinäkyvyyden ja algoritmisen vinouman välinen suhde: Opiskelijat ymmärtävät läpinäkyvyyden ja algoritmisen vinouman välisen yhteyden, tunnistavat läpinäkymättömyyden vaarat ja ymmärtävät, miten läpinäkyvyyden lisääminen voi auttaa tunnistamaan, ehkäisemään ja lieventämään tekoälyjärjestelmissä esiintyviä vinoumia. CUK4.3. Strategiat tekoälyjärjestelmien avoimuuden edistämiseksi: Opiskelijat tutustuvat yksinkertaisiin strategioihin, joilla voidaan edistää tekoälyjärjestelmien avoimuutta, kuten tulkittavien mallien käyttäminen, selkeä dokumentointi ja tekoälysovellusten päätöksentekoprosesseista tiedottaminen.	CUS4.1. Selittää avoimuuden käsite ja sen merkitys tekoälyjärjestelmien yhteydessä. CUS4.2. Tunnistaa avoimien tekoälyjärjestelmien edut sidosryhmille, mukaan lukien ymmärrettävyys, selitettävyys ja saavutettavuus. CUS4.3. Analysoida tekoälyjärjestelmien avoimuuden ja algoritmisen vinouman välistä suhdetta. CUS4.4. Analysoida tekoälyjärjestelmien läpinäkymättömyyden vaaroja ja tunnistaa, miten läpinäkyvyyden käsite voi auttaa ehkäisemään ja lieventämään tekoälyjärjestelmien vinoutuneita tuloksia. CUS4.5. Tunnistaa ja arvioida strategioita tekoälyjärjestelmien avoimuuden edistämiseksi, mukaan lukien tulkittavissa olevat mallit, selkeä dokumentointi ja päätöksentekoprosessien tehokas viestintä. CUS4.6. Soveltaa näitä strategioita tekoälyn kehittämisessä ja käyttää niitä avoimuuden lisäämiseksi ja algoritmisen vinoutuneisuuden lieventämiseksi.	CUA4.1. Arvostaa tekoälyjärjestelmien avoimuuden merkitystä luottamuksen rakentamisessa ja sidosryhmien ymmärtämisen mahdollistamisessa. CUA4.2. Arvostaa avoimuuden merkitystä tekoälyn eettisen kehityksen ja käytön edistämisessä. CUA4.3. Tunnustaa tekoälyjärjestelmien läpinäkymättömyyden vaarat ja avoimuuden merkityksen algoritmisen vinouman käsittelyssä ja lieventämisessä. CUA4.4. Kehittää sitoutumista tekoälyjärjestelmien läpinäkyvyyden lisäämiseen, jotta voidaan minimoida vääristyneet tulokset. CUA4.5. Kannattaa sellaisten strategioiden hyväksymistä ja täytäntöönpanoa, joilla edistetään tekoälyjärjestelmien avoimuutta. CUA4.6. Arvostaa tulkittavissa olevien mallien, selkeän dokumentoinnin ja tehokkaan viestinnän merkitystä avoimuuden kulttuurin edistämisessä ja algoritmisen vinouman lieventämisessä.

OSAAMISTULOSTEN KUVAUS

Oppijat saavat tietoa tekoälyjärjestelmien läpinäkyvyyden merkityksestä keskittyen peruskäsitteiden ymmärtämiseen, läpinäkyvyyden ja algoritmisen vinouman väliseen suhteeseen sekä strategioiden merkitykseen sen varmistamiseksi, että tekoälyjärjestelmät ovat ymmärrettäviä, selitettävissä ja sidosryhmien saatavilla. He tunnistavat reaali maailman vaikutukset ja ymmärtävät, miten tärkeitä tulkittavat mallit, selkeä dokumentaatio ja tehokas viestintä voivat olla läpinäkyvyyden kulttuurin edistämisessä ja algoritmisen vinouman lieventämisessä.

Kurssin osaamistavoitteet ovat seuraavat:

CU4.1. Avoimuuden merkitys tekoälyjärjestelmissä: Oppijat tunnistavat tekoälyn läpinäkyvyyden peruskäsitteen ja sen merkityksen sen varmistamisessa, että tekoälyjärjestelmät ovat ymmärrettäviä, selitettävissä ja sidosryhmien saatavilla. He tunnistavat avoimien tekoälyjärjestelmien hyödyt ja arvostavat niiden merkitystä luottamuksen rakentamisessa ja sidosryhmien ymmärtämisen mahdollistamisessa. Esimerkki: Vaikka tekoälymalli, joka on suunniteltu havaitsemaan syöpä, olisi vain 1 % väärässä, se voisi uhata ihmishenkiä. Tällaisissa tapauksissa tekoälyn ja ihmisten on tehtävä yhteistyötä, ja tehtävä helpottuu huomattavasti, kun tekoälymalli voi selittää, miten se on päätenyt tiettyyn päätökseen. Avoimuus tekee tekoälystä tiimipelaajan.

CU4.2. Läpinäkyvyyden ja algoritmisen vinouman välinen suhde: Oppijat löytävät yhteyden läpinäkyvyyden ja algoritmisen vinouman välille ja tunnistavat läpinäkyvyyden vaarat sekä sen, miten läpinäkyvyyden lisääminen voi auttaa tunnistamaan, estämään ja lieventämään tekoälyjärjestelmien vinoutuneita tuloksia. Oppijat tunnistavat avoimuuden merkityksen algoritmisen vinouman käsittelyssä ja lieventämisessä. Esimerkki: Usein tekoälyalgoritmit ovat vaikeaselkoisia siinä mielessä, että selitykset eivät ole kaikkien sidosryhmien saatavilla. Läpinäkyvyydellä voi olla erilaisia lähteitä. Joskus instituutiot tai yritykset eivät tiedota, milloin ne luottavat tekoälyjärjestelmiin tai miten nämä järjestelmät toimivat.

CU4.3. Strategiat tekoälyjärjestelmien avoimuuden edistämiseksi. Oppijat tutustuvat yksinkertaisiin strategioihin, joilla voidaan edistää tekoälyjärjestelmien avoimuutta, kuten tulkinnanvaraisten mallien käyttö, selkeä dokumentointi ja tekoälysovellusten päätöksentekoprosessien kommunikointi. Oppijat ymmärtävät, miten tärkeitä nämä strategiat ovat avoimuuden kulttuurin edistämässä ja algoritmisen vinouman lieventämisessä. Esimerkki: Tekoäly voi vaikuttaa eri sidosryhmiin, kuten käyttäjiin, asiakkaisiin, työntekijöihin, johtajiin, sääntelyviranomaisiin tai yhteiskuntaan. Avoimuuden ja vastuullisuuden varmistamiseksi sinun on otettava tekoälyn sidosryhmät mukaan ja annettava niille valtuudet koko tekoälyn elinkaaren ajan.

CU5 - Ihmisoikeudet ja oikeudenmukaisuus EQF4

OPPIMISTULOKSET

Pelillistetyn koulutuskokemuksen suoritettuaan oppija osaa seuraavat asiat		
TIEDOT	TAIDOT	ASENTEET
Teoreettinen/tosiasiallinen tietämys seuraavista: CUK5.1. Ihmisoikeuksien ja oikeudenmukaisuuden merkitys tekoälyjärjestelmissä: Oppijat oppivat ihmisoikeuksien ja oikeudenmukaisuuden merkityksen tekoälyjärjestelmien kehittämisessä ja käyttöönotossa ja tunnistavat, miten ne edistävät oikeudenmukaisia tuloksia ja ehkäisevät yksilöille ja yhteisöille aiheutuvaa haittaa. CUK5.2. Algoritmissen vinouman ja ihmisoikeuksien leikkauspiste: Oppijat tunnistavat algoritmissen vinouman ja ihmisoikeuksien välisen yhteyden ja tunnistavat, miten vinoutuneet tekoälyjärjestelmät voivat vaikuttaa suhteettomasti haavoittuvassa asemassa oleviin väestöryhmiin ja ylläpitää olemassa olevaa eriarvoisuutta. CUK5.3. Oikeudenmukaisuuden periaatteet tekoälyjärjestelmissä: Oppijat ymmärtävät tekoälyjärjestelmien oikeudenmukaisuuden perusperiaatteet, kuten yhtäläiset mahdollisuudet, syrjimättömyys, menettelyjen oikeudenmukaisuus, tasapuolisuus ja oikeudenmukaisuus, ja tunnistavat niiden roolin tulevien sukupolvien kannalta oikeudenmukaisten tulosten edistämässä.	CU5.1. Selittää ihmisoikeuksien ja oikeudenmukaisuuden merkityksen tekoälyjärjestelmissä ja niiden roolin oikeudenmukaisten tulosten edistämässä ja haittojen ehkäisemisessä. CU5.2. Soveltaa ihmisoikeuksia ja oikeudenmukaisuutta koskevia periaatteita tekoälyn kehittämisessä ja käyttöönotossa. CU5.3. Analysoida algoritmisen vinouman ja ihmisoikeuksien välistä suhdetta ja tunnistaa mahdolliset vaikuttavista haavoittuvassa asemassa oleviin väestöryhmiin ja epätasa-arvon ylläpitämiseen. CU5.4. Tunnistaa todellisia esimerkkejä ihmisoikeuksiin vaikuttavista tekoälyjärjestelmistä ja kehittää strategioita näiden ongelmien ratkaisemiseksi. CU5.5. Määritellä ja erottaa toisistaan erilaiset oikeudenmukaisuuden periaatteet tekoälyjärjestelmissä, kuten yhtäläiset mahdollisuudet, syrjimättömyys, menettelyllinen oikeudenmukaisuus, tasapuolisuus ja oikeudenmukaisuus. CU5.6. Soveltaa oikeudenmukaisuuden periaatteita tekoälyn kehittämisessä ja käyttää niitä edistämään oikeudenmukaisia tuloksia ja lieventämään algoritmien vinoumia tulevien sukupolvien puolesta.	CUA5.1. Arvostaa ihmisoikeuksien ja oikeudenmukaisuuden merkitystä tekoälyjärjestelmissä oikeudenmukaisuuden edistämiseksi ja haittojen ehkäisemiseksi. CUA5.2. Sitoutuu ihmisoikeuksien ja oikeudenmukaisuuden kunnioittamiseen tekoälyn kehittämisessä ja käyttöönotossa. CUA5.3. Arvostaa algoritmisen vinouman ja ihmisoikeuksien risteyskohdan merkitystä muokattaessa tekoälyjärjestelmien vaikutusta yksilöihin ja yhteisöihin. CUA5.4. Kehittää vastuuntuntoa käsitellä ja lieventää tekoälyjärjestelmien vinoutumien vaikutuksia ihmisoikeuksiin ja haavoittuvassa asemassa oleviin väestöryhmiin. CUA5.5. Tunnustaa tekoälyjärjestelmissä sovellettavien erilaisten reiluuden, tasapuolisuuden ja oikeudenmukaisuuden periaatteiden arvo algoritmisen vinouman lieventämisessä. CUA5.6. Edistää sitoutumista oikeudenmukaisuusperiaatteiden sisällyttämiseen tekoälyn kehittämiseen ja käyttöön, jotta saavutetaan oikeudenmukaisempia tuloksia tulevien sukupolvien etuja kunnioittaen.

OSAAMISTULOSTEN KUVAUS

Oppijat saavat tietoa ihmisoikeuksien ja oikeudenmukaisuuden merkityksestä tekoälyjärjestelmissä keskittymällä peruskäsitteiden ymmärtämiseen, algoritmisen vinouman ja ihmisoikeuksien sekä tekoälyjärjestelmien oikeudenmukaisuuden periaatteiden risteämiseen, reaali maailman vaikutusten tunnistamiseen, tasapuolisuuden ja oikeudenmukaisuuden eri periaatteiden arvon arvostamiseen tekoälyjärjestelmissä algoritmisen

vinouman pienentämiseksi ja oikeudenmukaisempien lopputulosten aikaansaamiseksi tulevien sukupolvien etua kunnioittaen.

Tämän kurssin osaamistavoitteet ovat seuraavat:

5.1. Ihmisoikeuksien ja oikeudenmukaisuuden merkitys tekoälyjärjestelmissä: Oppijat oppivat ihmisoikeuksien ja oikeudenmukaisuuden merkityksen tekoälyjärjestelmien kehittämisessä ja käyttöönotossa oikeudenmukaisten tulosten edistämiseksi ja haittojen ehkäisemiseksi. Oppijat ymmärtävät, että tekoäly tarjoaa kauaskantoisia hyötyjä inhimilliselle kehitykselle, mutta siihen liittyy myös riskejä (kuten muun muassa etuoikeutettujen ja muiden kuin etuoikeutettujen välisen jakautumisen lisääntyminen, yksilönvapauksien rapautuminen valvonnan kautta ja itsenäisen ajattelun ja harkinnan korvaaminen automaattisella hallinnalla).

5.2. Ymmärtää ihmisoikeuksien ja algoritmisen oikeudenmukaisuuden yhtymäkohdat: Tässä perusyksikössä oppijat tutustuvat ihmisoikeuksien ja algoritmisen oikeudenmukaisuuden kriittiseen risteyskohtaan. Teknologian moraalisia ja eettisiä vaikutuksia korostaneen Latourin (2005) teorioiden pohjalta oppijat tutkivat algoritmisten prosessien roolia ja vaikutusta ihmisoikeuksiin. Oppijat tutustuvat ihmisoikeuksien historialliseen kehitykseen ja siihen, miten se risteää algoritmisten teknologioiden syntymisen kanssa. Teoreettiset puitteet: Johdatus teoreettisiin kehyksiin, joissa analysoidaan ihmisoikeuksien ja algoritmisen oikeudenmukaisuuden välistä vuorovaikutusta. Esimerkki: Analysoidaan tapaustutkimuksia, joissa algoritmiset prosessit ovat olleet ristiriidassa ihmisoikeuksien kanssa, ja niihin liittyviä keskusteluja.

5.3. Oikeudenmukaisuuden periaatteet tekoälyjärjestelmissä: Oppijat tutustuvat tekoälyjärjestelmien oikeudenmukaisuuden perusperiaatteisiin, kuten yhtäläisiin mahdollisuuksiin, syrjimättömyyteen, menettelylliseen oikeudenmukaisuuteen, tasapuolisuuteen ja oikeudenmukaisuuteen, ja tunnistavat niiden arvon edistettäessä oikeudenmukaisia tuloksia tulevien sukupolvien puolesta.

CU6 - Tekoälyn etiikka, käytännön lähestymistapa EQF4

OPPIMISTULOKSET

TIEDOT	Pelillistetyn koulutuskokemuksen suoritettuaan oppija osaa seuraavat asiat	ASENTEET
Teoreettinen/tosiasiallinen tietämys seuraavilla aloilla: CUK6.1. Tekoälyn etiikan merkitys käytännössä: Oppijat oppivat, mikä merkitys on eettisten näkökohtien sisällyttämisellä tekoälyjärjestelmien kehittämiseen ja käyttöönottoon, ja ymmärtävät, miten etiikka voi lieventää algoritmisen vinouman kielteisiä seurauksia ja edistää luottamusta tekoälysovelluksiin. CUK6.2. Tekoälyjärjestelmien eettiset puitteet ja ohjeet: Oppijat tutustuvat erilaisiin tekoälyjärjestelmiin sovellettaviin eettisiin kehyksiin ja ohjeisiin, kuten tekoälyn eettisiin periaatteisiin, vastuullisiin tekoälykäytäntöihin ja toimialakohtaisiin säännöksiin, ja ymmärtävät niiden roolin eettisen tekoälyn kehittämisen ohjaamisessa. CUK6.3. Tekoälyn etiikan reaalimaailman sovellukset: Oppijat tutkivat käytännön esimerkkejä ja tapaustutkimuksia tekoälyjärjestelmistä eri aloilla, analysoivat eettisiä haasteita ja näkökohtia kussakin skenaariossa ja tunnistavat parhaat käytännöt eettisen tekoälykehityksen ja -käyttöönoton varmistamiseksi.	CUS6.1. Selittää eettisyyden merkitys tekoälyn kehittämisen ja käyttöönoton yhteydessä, mukaan luettuna eettisyyden rooli algoritmisen puolueellisuuden lieventämisessä ja luottamuksen edistämässä. CUS6.2. Soveltaa eettisiä näkökohtia tekoälyn kehittämiseen ja käyttöön vastuullisten ja luotettavien tekoälyjärjestelmien varmistamiseksi. CUS6.3. Tunnistaa ja analysoida erilaisia tekoälyjärjestelmiä koskevia eettisiä kehyksiä ja ohjeita, mukaan lukien tekoälyn eettiset periaatteet, vastuulliset tekoälykäytännöt ja toimialakohtaiset säädökset. CUS 6.4. Soveltaa tekoälyn kehittämisessä eettisiä viitekehyksiä ja ohjeita ja käyttää niitä vastuullisten ja eettisten tekoälykäytäntöjen ohjaamiseen. CUS6.5. Analysoida tekoälyjärjestelmiä koskevia todellisia esimerkkejä ja tapaustutkimuksia eettisten haasteiden ja näkökohtien tunnistamiseksi eri aloilla. CUS6.6. Soveltaa parhaita käytäntöjä ja eettisiä ohjeita tekoälyjärjestelmiä koskeviin reaalimaailman skenaarioihin eettisen kehityksen ja käyttöönoton varmistamiseksi.	CUA6.1. Arvostaa eettisyyden merkitystä tekoälyn kehityksessä ja käyttöönotossa algoritmisen vinoutuman lieventämiseksi ja luottamuksen lisäämiseksi. CUA6.2. Sitoutuminen eettisten näkökohtien sisällyttämiseen tekoälyjärjestelmiin vastuullisten ja luotettavien tekoälysovellusten varmistamiseksi. CUA6.3. Arvostaa eettisten kehysten ja ohjeiden merkitystä tekoälyn eettisen kehittämisen ja käytön ohjaamisessa. CUA6.4. Kasvattaa sitoutumista eettisten kehysten ja ohjeiden soveltamiseen vastuullisten ja eettisten tekoälykäytäntöjen varmistamiseksi. CUA6.5. Tunnustaa, miten tärkeää on ymmärtää ja käsitellä eettisiä haasteita todellisissa tekoälysovelluksissa. CUA6.6. Edistää sitoutumista parhaiden käytäntöjen ja eettisten ohjeiden omaksumiseen reaalimaailman skenaarioissa, jotta voidaan varmistaa tekoälyn eettinen kehittäminen ja käyttöönotto.

OSAAMISTULOSTEN KUVAUS

Tavoite: Tämä yksikkö on suunniteltu EQF4-tason oppijoille, jotta he voivat syventyä tekoälyn etiikan käytännön näkökohtiin ja sisällyttää tekoälyn vastuullisen kehittämisen ja käytön periaatteet reaalimaailman skenaarioihin. Oppijoita kannustetaan toimimaan ennakoivasti eettisten ohjeiden täytäntöönpanossa tekoäly-ympäristöissä, hyödyntämään teoreettista tietoa ja muuntamaan se toimiviksi oivalluksiksi.

6.1. Tekoälyn eettisten vaikutusten ymmärtäminen

Tässä ensimmäisessä moduulissa oppijat saavat aluksi vankan perustan tekoälyn eri eettisistä ulottuvuuksista. Floridin ym. (2018) antamien, tekoälyyn liittyviä eettisiä huolenaiheita koskevien näkemysten mukaisesti oppijat

tutustuvat tekoälyn mahdollisiin vaikutuksiin yhteiskuntaan, yksilöihin ja globaaliin yhteisöön.

Selitys:

Eettinen perusta: Erilaisten eettisten teorioiden ja niiden vaikutusten analysointi ja ymmärtäminen tekoälyyn.

Moraaliset vaikutukset: Yksityiskohtainen tutkimus tekoälyn ja automatisoitujen päätöksentekojärjestelmien moraalisisista vaikutuksista.

Esimerkki: Analyysi todellisista tapaustutkimuksista, jotka kuvaavat tekoälyteollisuuden eettisiä ongelmia.

6.2. Tekoälyn epäeettisten käytäntöjen tunnistaminen ja lieventäminen

Tämä yksikkö käsittelee tarvetta tunnistaa ja lieventää mahdollisia epäeettisiä käytäntöjä tekoälyn alalla. Oppijat oppivat Bostromin (2014) pohdinnoista, joissa hän tutki tekoälyn tulevaisuutta ja sen yhteensovittamista inhimillisten arvojen kanssa, miten tunnistaa ja estää epäeettiset käytännöt tekoälyn kehittämisessä ja käyttöönotossa.

Selitys:

Epäeettisten käytäntöjen tunnistaminen: Tutkimus erilaisista indikaattoreista, jotka merkitsevät epäeettisiä käytäntöjä tekoälyn alalla.

Lieventämisstrategiat: Strategioiden kehittäminen tekoälysovelluksista aiheutuvien mahdollisten epäeettisten seurausten lieventämiseksi.

Esimerkki: Roolileikit, joissa opiskelijat tunnistavat ja käsittelevät epäeettisiä käytäntöjä simuloidussa tekoäly-ympäristössä.

6.3. Tekoälyn eettisten ohjeiden käytännön soveltaminen

Tämän moduulin oppijoita kannustetaan muuntamaan teoreettinen ymmärrys käytännön toiminnaksi. Ryan & Stahlin (2020) teosten pohjalta, jotka keskittyvät eettisten ohjeiden laatimiseen, opiskelijat osallistuvat tehtäviin, jotka edistävät näiden ohjeiden käytännön soveltamista tekoälyskenaarioissa.

Selitys:

Eettisten suuntaviivojen kehittäminen: Eettisten ohjeiden laatiminen, joita voidaan soveltaa todellisissa tekoälyhankkeissa.

Toteutus todellisissa tilanteissa: Työpajat ja toimet, joissa keskitytään kehitettyjen suuntaviivojen täytäntöönpanoon todellisissa tekoälyhankkeissa.

Esimerkki: Oppijat osallistuvat projektiin, jossa he kehittävät ja soveltavat eettisiä ohjeita todellisessa tekoälyprojektissa.

6.4. Vastuullisen tekoälyn kehittämisen ja käyttöönoton edistäminen

Viimeisessä yksikössä keskitytään sellaisen ympäristön edistämiseen, joka kannustaa vastuulliseen tekoälyn kehittämiseen ja käyttöönottoon. Jobin ym. (2019) ehdottaman kehyksen pohjalta, jossa käsiteltiin globaaleja näkökulmia tekoälyn etiikkaan, oppijat tutkivat menetelmiä, joilla edistetään globaalia yhteistyötä ja vastuullisen tekoälyn luomista.

Selitys:

Maailmanlaajuinen yhteistyö: Globaalin yhteistyön merkityksen ymmärtäminen vastuullisen tekoälyn edistämisessä.

Yhteisön sitoutuminen: Kannustetaan yhteisöä sitoutumaan ja osallistumaan tekoälyn eettiseen kehittämiseen.

Esimerkki: Oppijat kehittävät strategioita ja suunnitelmia yhteisön sitoutumisen ja maailmanlaajuisen yhteistyön edistämiseksi tekoälyn etiikan alalla.

4.3. CHARLIE WP2:n oppimistulokset aikuisille ja nuorille EQF2 (SERIOUS GAME).

Algoritmisen vinouman pohdintapeli on opettavainen ja vuorovaikutteinen peli, joka johdattaa EQF2-tason osallistujat algoritmiin liittyvien vinoumien tärkeyteen aiheeseen. Pelin aikana tutkitaan algoritmien peruskäsitteitä, niiden mahdollisia vinoumia ja vaikutuksia, joita näillä ennakkoluuloilla voi olla yhteiskunnan eri osa-alueisiin.

Algoritmisen vinouman purkaminen -kurssin (UAB) ensisijaisena tavoitteena on tarjota oppijoille kiinnostava oppimisympäristö, jossa he voivat tutkia aihetta interaktiivisten toimintojen ja skenaarioiden avulla. Esittämällä sisältö helposti lähestyttävässä ja miellyttävässä muodossa odotetaan, että osallistujat ymmärtävät ja tiedostavat syvällisemmin algoritmisen vinouman merkityksen.

Työpaketin odotetut tulokset:

- Tutustuttaa osallistujat algoritmisen vinouman käsitteeseen ja sen merkitykseen nykymaailmassa sekä korostaa, miten tärkeää on käsitellä tätä kysymystä eri aloilla.
- Tutustuttaa oppijat tosielämän esimerkkeihin algoritmisesta vinoumasta ja havainnollistaa vinoutuneisiin algoritmeihin liittyviä mahdollisia seurauksia ja eettisiä vaikutuksia.
- Edistää osallistujien kriittisen ajattelun taitoja, jotta he voivat analysoida ja arvioida algoritmeja ja niiden mahdollisia vinoumia eri yhteyksissä.
- Kehittää oppijoiden ymmärrystä erityyppisistä algoritmisista vinoumista, mukaan lukien dataan perustuva, malliin perustuva ja ihmislähtöinen vinouma.
- Kannustaa osallistujia tutkimaan oikeudenmukaisuuden, avoimuuden ja vastuuvollisuuden merkitystä tekoälyjärjestelmien kehittämisessä ja käyttöönotossa.
- Lisätä osallistujien tietoisuutta tekoälyn kehittämistä ohjaavista eettisistä periaatteista ja ohjeista, kuten non-maleficence-periaate, ihmisoikeudet ja yksityisyys.
- Haastaa osallistujat soveltamaan uutta tietämystään tunnistamaan mahdollisia vääristymiä simuloiduissa tekoälysovelluksissa ja ehdottamaan ratkaisuja näiden vääristymien lieventämiseksi.
- Edistää oppijoiden välistä yhteistyötä ja viestintää, kun he työskentelevät yhdessä ratkaistakseen algoritmisiin vinoumiin liittyviä ongelmia ja jakaakseen näkemyksiään.

- Osallistuminen pohdintatehtäviin, joissa osallistujia kannustetaan pohtimaan omia näkökulmiaan ja ennakkoluulojaan sekä sitä, miten ne voivat vaikuttaa tekoälyjärjestelmien käyttöön.
- Innostaa osallistujia toimimaan ennakoivasti eettisen tekoälyn kehittämisen ja käyttöönoton puolesta omissa yhteisöissään ja ammatillisissa ympäristöissään ja edistämään vastuuntuntoa ja kansalaisvaikuttamista.

Pelin päätyttyä EQF2-tason oppijat ovat saaneet perusymmärryksen algoritmisesta vinoumasta ja sen mahdollisista seurauksista. He kehittävät ensimmäisiä ongelmanratkaisutaitoja, joiden avulla he voivat tunnistaa tekoälysovellusten vääristymät. Lisäksi osallistujat ovat parantaneet viestintä- ja tiimityöskentelytaitojaan ja edistäneet samalla vastuuntuntoa ja tietoisuutta tekoälyn ja teknologian käytön eettisistä näkökohdista.

4.3.1. Algoritmisen vinouman pohdintapelin sisältö

EQF2-tason oppijoille suunnatun pelin sisältö voidaan jakaa viiteen moduuliin:

1. Johdanto algoritmiseen vinoumaan:

Algoritmien ja tekoälyn peruskäsitteet

Johdatus tekoälyjärjestelmien vinoumaan ja oikeudenmukaisuuteen

Todellisia esimerkkejä algoritmisesta puolueellisuudesta

2. Algoritmien vinoumien tyypit:

Dataan perustuva vinouma

Malliin perustuva vinouma

Ihmislähtöinen vinouma

Esimerkkejä ja skenaarioita kustakin vinouman tyypistä

3. Algoritmien vinoumien tunnistaminen ja siihen puuttuminen:

Strategiat tekoälysovellusten vinoumien havaitsemiseksi

Perusmenetelmät vinoumien lieventämiseksi

Yksilöiden ja organisaatioiden rooli algoritmiseen vinoumaan puuttumisessa

4. Tekoälyn eettiset näkökohdat:

Eettisyyden merkitys tekoälyn kehittämisessä ja käytössä

Johdatus eettisiin periaatteisiin, kuten non-maleficence-periaate, vastuuvollisuus ja avoimuus.

Ihmisoikeuksien, oikeudenmukaisuuden ja tekoälyn etiikan välinen yhteys

5. Pelitoiminnot ja haasteet:

Vuorovaikutteiset skenaariot, joissa pelaajat tunnistavat tekoälysovellusten vinoumia ja puuttuvat niihin.

Eettiseen tekoälyn kehittämiseen liittyvät tiimipohjaiset ongelmanratkaisutehtävät

Pohdinta- ja keskustelutilaisuudet, joilla vahvistetaan oppimista ja edistetään tietoisuutta algoritmisista vinoumista.

Koko pelin ajan oppijat osallistuvat yhteistoiminnallisiin tehtäviin ja haasteisiin, joiden tarkoituksena on vahvistaa käsitteiden ymmärtämistä ja edistää samalla tiimityötä, viestintätaitoja ja ongelmanratkaisutaitoja.

4.3.2. Kohderyhmän kuvaus, EQF2-ominaisuudet ja piirteet

Eurooppalaisen tutkintojen viitekehyksen (EQF) taso 2 (EQF2) edustaa perustavaa laatua olevaa osaamistasoa eri oppiaineissa tai taidoissa (Euroopan komissio, 2023). EQF2-tason osallistujat ovat tyypillisesti oppimispolkunsa alkuvaiheessa tai suorittavat ammatillista peruskoulutusta. EQF2-tason tärkeimpiä piirteitä ja ominaisuuksia ovat muun muassa seuraavat:

- Perustiedot: EQF2:n osallistujilla on perustavanlaatuinen ymmärrys valitsemansa aihealueen keskeisistä käsitteistä, periaatteista ja ajatuksista syventymättä monimutkaisiin teorioihin tai syventäviin aiheisiin.
- Käytännön taidot: EQF2-tason osallistujat osaavat soveltaa perustietojaan ja suorittaa alaansa liittyviä yksinkertaisia tehtäviä ja toimintoja käyttäen perustyökaluja ja -tekniikoita valvotusti tai ohjatusti.
- Kognitiiviset taidot: EQF2-oppijat osaavat käyttää kognitiivisia perustaitoja, kuten ongelmanratkaisua ja päätöksentekoa, tutuissa ja yksinkertaisissa tilanteissa. He osaavat noudattaa yksinkertaisia ohjeita ja menettelytapoja tehtävien suorittamiseksi.
- Viestintä: EQF2:n oppijat pystyvät välittämään perustiedot ja -ajatuksia selkeästi ja tiiviisti sekä suullisesti että kirjallisesti tutuissa yhteyksissä.
- Yhteistyö ja tiimityö: EQF2-tason oppijat pystyvät työskentelemään tehokkaasti muiden kanssa ryhmässä tai tiimissä, osoittamaan yhteistyökykyä ja ymmärtämään ihmissuhteiden perusdynamikkaa.
- Vastuu ja autonomia: EQF2-oppijat voivat ottaa vastuun toimistaan ja päätöksistään rajallisten tietojensa ja taitojensa puitteissa. He osaavat työskennellä valvotusti, noudattaa ohjeita ja pyytää tarvittaessa apua.

- Sopeutumiskyky ja joustavuus: EQF2-tason oppijat pystyvät sopeutumaan yksinkertaisiin muutoksiin oppimis- tai työympäristössään ja selviytymään perushaasteista tai takaiskuista ohjauksen ja tuen avulla.
- Itsetuntemus ja itsereflektio: EQF2-oppijat voivat pohtia omia oppimiskokemuksiaan ja tunnistaa parannuskohteita, mikä osoittaa perustason itsetuntemusta ja sitoutumista henkilökohtaiseen kasvuun.
- Eettinen tietoisuus: EQF2-tasolla oppijat osaavat tunnistaa aihealueeseensa liittyvät eettiset perusperiaatteet ja -ohjeet ja osoittaa ymmärtävänsä niiden merkityksen jokapäiväisissä tilanteissa.
- Elinikäinen oppiminen: EQF2-oppijoita kannustetaan kehittämään kiinnostusta valitsemaansa alaa kohtaan ja kehittämään elinikäisen oppimisen ajattelutapaa, edistämään uteliaisuutta ja halua jatkaa koulutusta ja taitojen kehittämistä EQF2-tason jälkeen.

4.3.3. Yleinen kompetenssimatriisi Algoritmisen vinouman pohdintapelille EQF2:ssa

OPPIMISTULOKSET		
	Pelin päätyttyä oppija osaa seuraavat asiat	
TIEDOT	TAIDOT	ASENTEET
1. Algoritmien peruskäsitteen ymmärtäminen ja niiden käyttö jokapäiväisessä elämässä, mukaan lukien yksinkertaiset esimerkit niiden sovelluksista eri aloilla. 2. Tunnistaa oikeudenmukaisuuden ja puolueettoman päätöksenteon merkityksen algoritmisissa prosesseissa. 3. Ymmärtää algoritmisen vinouman peruskäsitteen, mukaan lukien ymmärrys siitä, miten tekoälyjärjestelmät voivat sisältää vinoumia ja niiden mahdollisia seurauksia. 4. Tunnistetaan yksinkertaisia esimerkkejä algoritmisesta vinoumasta reaali maailman tilanteissa ja korostetaan, miten tärkeää on puuttua näihin vinoumiin tasapuolisten tulosten varmistamiseksi. 5. Ymmärtää datan roolin tekoälyjärjestelmien	1. Analysoida yksinkertaisia tosielämän skenaarioita algoritmisen vinouman ja sen mahdollisten vaikutusten tunnistamiseksi. 2. Soveltaa kriittisen ajattelun perustaitoja arvioidakseen tekoälyjärjestelmien ja niiden päätöksentekoprosessien oikeudenmukaisuutta. 3. Kehittää yksinkertaisia strategioita, joilla voidaan minimoida tekoälyjärjestelmissä esiintyvää vääristymää, esimerkiksi käyttämällä monipuolisempaa ja edustavampaa dataa. 4. Viestiä tehokkaasti algoritmisen vinouman peruskäsitteistä ja sen seurauksista vertaisryhmille ja muille kuin erikoistuneille kohderyhmille. 5. Käyttää ongelmanratkaisutaitoja tekoälyjärjestelmien algoritmiseen vinoumaan ja	1. Oikeudenmukaisuuden ja puolueettoman päätöksenteon merkityksen arvostaminen tekoälyjärjestelmissä ja algoritmisen vinoutumisen mahdollisten seurausten tunnistaminen yksilöille ja yhteiskunnalle. 2. Eettinen vastuu, jossa tunnustetaan yksilöiden rooli tekoälyjärjestelmien kehittämisessä ja käyttöönotossa sekä tarve toimia eettisesti ja vastuullisesti. 3. Avoimuus erilaisille näkökulmille, halukkuus ottaa huomioon erilaisia näkökulmia ja kokemuksia, kun käsitellään algoritmiseen vinoumaan ja oikeudenmukaisuuteen liittyviä kysymyksiä. 4. Sitoutuminen jatkuvaan oppimiseen ja henkilökohtaiseen kasvuun ja sen ymmärtäminen, että algoritmiseen vinoumaan puuttuminen edellyttää jatkuvaa oppimista ja sopeutumista. 5. Yhteistyö ja tiimityö, muiden panoksen arvostaminen ja sen

toiminnassa ja sen, miten vääristynyt tai epätäydellinen data voi johtaa vääristyneisiin päätöksiin. 6. Saada perustietämys tekoälyn eettisistä näkökohdista ja puolueettomien algoritmien kehittämisen ja käyttöönoton tärkeydestä. 7. Tunnistaa tekoälyjärjestelmien läpinäkyvyyden ja vastuullisuuden tarpeen, jotta voidaan varmistaa niiden vastuullinen käyttö ja puuttua mahdollisiin vinoumiin. 8. Ymmärtää ihmisoikeuksien ja oikeudenmukaisuuden merkityksen tekoälyn yhteydessä ja puolueettomien algoritmien roolin näiden periaatteiden edistämiseksi. 9. Kehittää perustietämystä algoritmisen vinouman mahdollisista oikeudellisista ja yhteiskunnallisista vaikutuksista ja asianmukaisten säännösten ja ohjeiden tarpeesta. 10. Kasvattaa kiinnostusta oppia lisää algoritmisesta vinoumasta, tekoälyn etiikasta ja niihin liittyvistä aiheista ja edistää sitoutumista elinikäiseen oppimiseen ja vastuulliseen sitoutumiseen tekoälyteknologiaan.	oikeudenmukaisuuteen liittyvien perushaasteiden ratkaisemiseen. 6. Pohtia henkilökohtaisia kokemuksia ja vinoumia ymmärtääkseen paremmin, miten tärkeää on edistää oikeudenmukaisuutta ja puolueetonta päätöksentekoa tekoälyjärjestelmissä. 7. Tehdä yhteistyötä muiden kanssa algoritmiseen vinoumaan liittyvien mahdollisten ongelmien tunnistamiseksi ja asianmukaisten ratkaisujen kehittämiseksi. 8. Osoittaa eettistä perustietämystä tekoälyyn liittyen ja ymmärtää vastuullisen tekoälyn kehittämisen ja käyttöönoton merkityksen. 9. Sopeutuu erilaisiin skenaarioihin, joihin liittyy algoritmisia vinoumia, ja tiedostaa jatkuvan oppimisen ja parantamisen tarpeen. 10. Osoittaa ennakoivaa asennetta algoritmiseen vinoumaan puuttumista kohtaan, ottaa aloitteellisesti oppia lisää ja osallistua asiaankuuluviin keskusteluihin ja toimiin.	tunnustaminen, että on tärkeää työskennellä yhdessä algoritmisen vinouman kaltaisten monimutkaisten kysymysten käsittelemiseksi. 6. Kriittinen ajattelu, oletusten kyseenalaistaminen ja halu arvioida uudelleen henkilökohtaisia uskomuksia ja ennakkoluuloja uusien tietojen tai näkökulmien valossa. 7. Empatia ja myötätunto, algoritmisen vinouman vaikutuksen huomioon ottaminen eri yksilöihin ja yhteisöihin sekä pyrkimys edistää oikeudenmukaisuutta ja tasapuolisuutta tekoälyjärjestelmissä. 8. Yksityisyyden ja tietosuojan kunnioittaminen, käyttäjien yksityisyyden säilyttämisen ja henkilötietojen suojaamisen tärkeyden tunnistaminen tekoälysovelluksissa. 9. Sosiaalinen tietoisuus, algoritmisen vinouman laajempien yhteiskunnallisten vaikutusten ymmärtäminen ja tarve yhteisiin toimiin näiden haasteiden ratkaisemiseksi. 10. Ennakoiva sitoutuminen, oma-aloitteinen osallistuminen algoritmiseen vinoumaan ja oikeudenmukaisuuteen liittyviin keskusteluihin ja toimiin sekä vastuullisen tekoälyn kehittämisen ja käyttöönoton puolustaminen.
--	---	--

Viitteet

- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. ProPublica, toukokuu, 23.
- Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Abstraction in Sociotechnical Systems. In ACM Conference on Fairness, Accountability, and Transparency (s. 59-68).
- Benjamin, R. (2019). Race After Technology: Abolitionist Tools for the New Jim Code. Polity.
- Bentley, J. (1999). Programming Pearls. Addison-Wesley.
- Billett, S. (2009). Realising the educational worth of integrating work experiences in higher education. Studies in Higher Education, 34(7), 827-843.
- Börner, K., Contractor, N., Falk-Krzesinski, H. J., Fiore, S. M., Hall, K. L., Keyton, J., ... & Uzzi, B. (2010). A multi-level systems perspective for the science of team science. Science Translational Medicine, 2(49), 49cm24.
- Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press.
- Cavoukian, A. (2010). Privacy by Design: The 7 Foundational Principles. Information and Privacy Commissioner of Ontario, Canada.
- Cedefop - European Centre for the Development of Vocational Training - European Qualifications Framework (EQF): <https://www.cedefop.europa.eu/en/events-and-projects/projects/european-qualifications-framework-eqf>.
- Cook, W. (2016). In Pursuit of the Traveling Salesman: Mathematics at the Limits of Computation. Princeton University Press.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). Introduction to Algorithms. MIT Press.
- Cottrell, S. (2018). Skills for success: Personal development and employability. Macmillan International Higher Education.
- Cuseo, J. (2010). The empirical case for the first-year seminar: Promoting positive student outcomes and campus-wide benefits. The First-Year Seminar: Research-Based Recommendations for Course Design, Delivery, and Assessment. Dubuque, IA: Kendall Hunt.

- Danks, D., & London, A. J. (2017). Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms. ACM Digital Library.
- Diaz, A., & Sonsini, G. (2016). Algorithmic decision-making and the law. Telecommunications Policy, 40(11), 1027-1040.
- Doe, J., & Roe, J. (2020). Understanding Algorithm Models: A Comprehensive Approach. Journal of Computer Science, 18(3), 300-316.
- Eraut, M. (2004). Informal learning in the workplace. Studies in Continuing Education, 26(2), 247-273.
- Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.
- European Commission - European Qualifications Framework (EQF): <https://ec.europa.eu/ploteus/en/content/descriptors-page>.
- European Commission - The European Higher Education Area (EHEA): https://ec.europa.eu/education/policies/higher-education/european-higher-education-area_en.
- European Commission (2019). Ethical Guidelines of Trustworthy AI. High-Level Expert Group on Artificial Intelligence. Saatavilla 20.6.2023 <https://op.europa.eu/o/opportal-service/download-handler?identifier=d3988569-0434-11ea-8c1f-01aa75ed71a1&format=pdf&language=en&productionSystem=cellar&part=>.
- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. & Srikumar, M. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Berkman Klein Center for Internet and Society. Harvardin yliopisto. Saatavilla 20.6.2023 https://dash.harvard.edu/bitstream/handle/1/42160420/HLS%20White%20Paper%20Final_v3.pdf?sequence=1&isAllowed=y.
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People-an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds and machines, 28, 689-707.
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Schafer, B. (2018). AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28(4), 689-707.

- Friedman, B. & Nissenbaum, H. (1996). Bias in Computer Systems. ACM Transactions on Information Systems, Vol. 14, No. 3, July 1996, Pages 330-347. Saatavilla 20.6.2023 osoitteessa <https://nissenbaum.tech.cornell.edu/papers/biasincomputers.pdf>.
- Jobin, A., Ienca, M. & Vayena, E. (2019). Tekoäly: The Global Landscape of Ethics Guidelines. Saatavilla 20.6.2023 <https://arxiv.org/ftp/arxiv/papers/1906/1906.11668.pdf>.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.
- Johnson, D. (2003). A theoretician's guide to the experimental analysis of algorithms. Data Structures, Near Neighbor Searches, and Methodology: Fifth and Sixth DIMACS Implementation Challenges, 59.
- Karp, R. (2010). The art of algorithm optimization. ACM Computing Surveys (CSUR), 42(4), 1-28.
- Knuth, D. E. (1997). The Art of Computer Programming, Volume 1: Fundamental Algorithms. Addison-Wesley.
- Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. Management science, 65(7), 2966-2981.
- Lange, A. R. & Duarte, N. (2017). Understanding Bias in Algorithmic Design. Saatavilla 20.6.2023 <https://medium.com/impact-engineered/understanding-bias-in-algorithmic-design-db9847103b6e>.
- Latour, B. (2005). Understanding Bias in Algorithmic Design Understanding Bias in Algorithmic Design. Oxford University Press.
- Lee, N. T., Resnick, P., & Barton, G. (2019). Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms. Saatavilla 20.6.2023 osoitteessa <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>.
- Li, Y., Hu, B., Chen, M., & Zhong, Q. (2017). Understanding the Limitations of Algorithms: A Comprehensive Review. Journal of Information Science, 43(4), 528-545.
- Mansfield, R. S. (2009). The competencies required by the role of human resource developers. Performance Improvement Quarterly, 22(2), 33-49.

- Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *Ai & Society*, 35, 957-967.
- Miller, G. (2022). Stakeholder Roles in Artificial Intelligence Projects. *Project Leadership and Society*, Vol. 3, joulukuu 2022. Saatavilla 20.6.2023 <https://www.sciencedirect.com/science/article/pii/S266672152200028X>.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 2053951716679679.
- Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 2053951716679679.
- Morley, J., Floridi, L., Kinsey, L. & Elhalal, A. (2019). From What to How: An Overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices. Saatavilla 20.6.2023 https://www.researchgate.net/profile/Jessica-Morley/publication/333103986_From_What_to_How_An_Overview_of_AI_Ethics_Tools_Methods_and_Research_to_Translate_Principles_into_Practices/links/5cdbb95ca6fdccc9ddae53db/From-What-to-How-An-Overview-of-AI-Ethics-Tools-Methods-and-Research-to-Translate-Principles-into-Practices.pdf?origin=publication_detail.
- Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.
- Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.
- Novelli, C., Taddeo, M. & Floridi, L. (2023). Accountability in Artificial Intelligence: What It Is and How It Works. *AI & SOCIETY*. Available at 20.6.2023 <https://link.springer.com/content/pdf/10.1007/s00146-023-01635-y.pdf?pdf=button>
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.

- Paraschakis, D. (2018). Algorithmic and ethical aspects of recommender systems in E-commerce (Doctoral dissertation, Malmö University, Faculty of Technology and Society).
- Parker, G., Van Alstyne, M., & Choudary, S. (2016). Platform Revolution: How Network
- Pasquale, F. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard University Press.
- Perrault R, Yoav S, Brynjolfsson E, Jack C, Etchmendi J, Grosz B, Terah L, James M, Saurabh M, Carlos NJ (2019). *Artificial Intelligence Index Report 2019*. https://wp.oecd.ai/app/uploads/2020/07/ai_index_2019_introduction.pdf
- Prates, M. O., Avelar, P. H., & Lamb, L. C. (2020). Assessing gender bias in machine translation: a case study with google translate. *Neural Computing and Applications*, 32, 6363-6381.
- Ryan, M., & Stahl, B. C. (2020). Artificial Intelligence Ethics Guidelines for Developers and Users: Clarifying Their Content and Normative Implications. *Journal of Information, Communication and Ethics in Society*.
- Stewart, J., & Bartrum, D. (1999). Towards a competency matrix for human resource development. *Industrial and Commercial Training*, 31(5), 176-181.
- UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence. Saatavilla 20.6.2023 https://unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf
- Wong, P. H. (2020). Democratizing algorithmic fairness. *Philosophy & Technology*, 33, 225-244.
- Zhou, J., Chen, Z., Berry, A., Reed, M., Zhang, S. & Savage. S. (2020). A Survey on Ethical Principles of AI and Implementations. IEEE Xplore. Saatavilla 20.6.2023 https://opus.lib.uts.edu.au/bitstream/10453/146673/2/IEEE_ETHAI2020_ethicalAI_Survey.pdf
- Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. PublicAffairs.