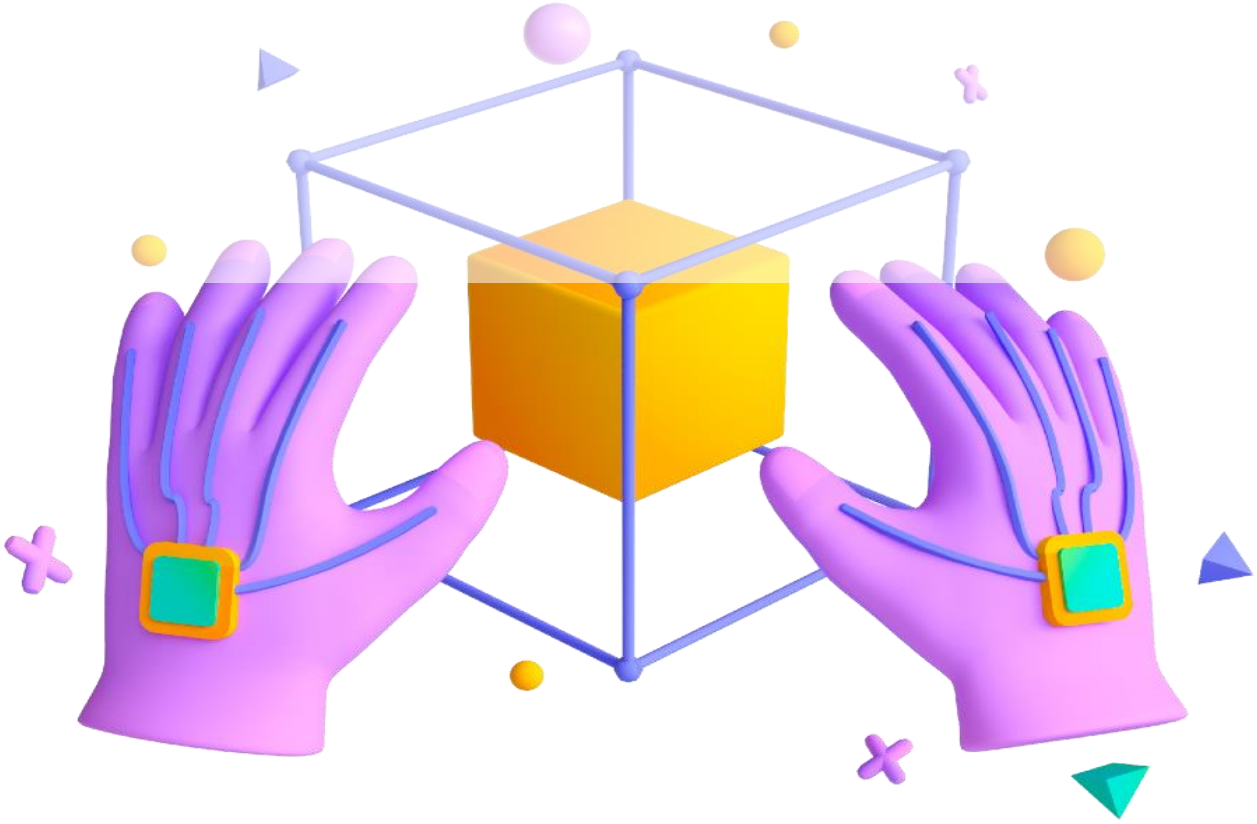
CHARLIE - HÅNDTERING AF BIAS I BIG DATA TIL AI OG MACHINE LEARNING

WP2 Kompetencematricer



Co-funded by the European Union

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

Dette projekt er udarbejdet med støtte fra Europa-Kommissionen. Publikationen afspejler udelukkende forfatterens synspunkter, og Kommissionen kan ikke holdes ansvarlig for nogen brug, der måtte blive gjort, af oplysningerne heri.

Indholdsfortegnelse

1. INTRODUKTION TIL CHARLIE-PROJEKTET

1.1. CHARLIE-projektet: et overblik

CHARLIE er et ERASMUS+ KA2-projekt med en implementeringsperiode på 30 måneder i perioden 30/12/2022 - 29/06/2025. Projektet gennemføres af et konsortium bestående af seks (6) partnere fra fem (5) europæiske lande: Spanien, Portugal, Rumænien, Finland og Danmark.

Kunstig intelligens (AI) er en del af vores hverdag: Lige fra den algoritme, der genkender vores ansigt, når vi går ned ad gaden, og som forsyner politiets biometriske sikkerhedstjenester med data, til den algoritme, der vælger de reklamer, vi ser på vores sociale medier – AI er overalt. Men selvom *machine learning* (ML) og AI er matematik, har disse ikke altid ret, fordi de data, der behandles for at nå frem til et resultat, kan være *biased*, dvs. partiske eller forudindtaget – og ofte er de det også. Samfundsvidenskaberne har studeret menneskelig bias i mange år. Dette skyldes de implicitte associationer, som afspejler de fordomme, vi ikke selv er bevidste om, og som kan resultere i yderligere negative resultater. AI og ML er ikke designet til at træffe etiske beslutninger; teknologierne indeholder ikke en algoritme for etik. De vil altid komme med forudsigelser baseret på, hvordan verden fungerer i dag, og derfor bidrager de til at fremme den forudindtagethed og diskriminerende praksis, der er systematisk forankret i vores samfund i dag. Med den udbredte brug af AI- og ML-teknologier, der ofte ejes af store tech-virksomheder, hvis eneste formål er at skabe profit, er der et presserende behov for en menneskecentreret tilgang til *tech* og bruge den tilgang for at løse sociale problemer i stedet for at bidrage til dem. Europa-Kommissionen fremlagde i meddelelser af 25.04.18 og 07.12.18 sin vision for kunstig intelligens, hvori de går ind for "etisk, sikker og banebrydende kunstig intelligens produceret i Europa". AI-systemer skal være menneskecentrerede og forpligtede på at blive anvendt i menneskehedens og almenvellets tjeneste med det mål at forbedre menneskers velfærd og frihed. Selvom AI-systemer giver store muligheder, indebærer de også visse risici, som skal håndteres på en hensigtsmæssig og rimelig måde. Vi har nu en række muligheder for at forme deres udvikling.

CHARLIE har til hensigt at sikre, at vi kan stole på de sociotekniske miljøer, som systemerne er indlejret i. Vi ønsker også, at producenter af AI-systemer får en konkurrencemæssig fordel ved at integrere troværdig AI i deres produkter og services. Det indebærer, at man søger at maksimere fordelene ved AI-systemer, samtidig med at man forebygger og minimerer deres risici. Videregående uddannelser (HE – *higher education*), voksenuddannelser (AE – *adult education*) samt udskolingen og ungdomsuddannelser (12-18 år – *Youth*)

har brug for nye og innovative undervisningsforløb, der kan imødekomme dette kompetencegab, og som kan udstyre eleverne med viden og kompetencer til at bidrage til en mere etisk tilgang til teknologisk udvikling. Behovet for at gøre tech-uddannelser mere menneskelige er i tråd med Handlingsplanen for Digital Uddannelse (*the Digital Education Action Plan*), der indeholder specifikke tiltag til at håndtere etiske implikationer og udfordringer ved at bruge kunstig intelligens og data i forskellige uddannelsesforløb.

1.2. Projektets formål

CHARLIE sigter mod at håndtere *bias* i big data, der bruges til AI og ML, ved at skabe en større bevidsthed om de negative konsekvenser af manglen på en kritisk og etisk tilgang til tech-uddannelse. CHARLIE hovedformål er at:

1. Øge de videregående uddannelsesinstitutioners evne til at give deres studerende online læringsmuligheder, der opfylder samfundets behov, men som også er skræddersyet til de studerendes læringsbehov;
2. Øge tech-studerendes sociale og etiske kompetencer, så de kan engagere sig positivt, kritisk og etisk i AI/ML-teknologi;
3. Udstyre lærere/professorer med digitale og engagerende tilgange til effektiv undervisning i emnet (især i onlineundervisning);
4. Skabe synergier mellem videregående uddannelsesinstitutioner, voksenundervisningsinstitutioner og udskoling- samt ungdomsuddannelsesinstitutionerne på regionalt niveau inden for uddannelse i AI-etik;
5. Styrke overførbarheden mellem akademiske kurser om AI-bias til voksenundervisning og erhvervsuddannelser;
6. Øge bevidstheden om emnet på samfundsniveau.

1.3. Forventede, overordnede resultater af CHARLIE-projektet for hver WP (*Work Package*)

1. Kompetencematricer for kurserne "Algoritmisk Bias" (EQF6), "Etisk AI Micro-Credential" (EQF4) og Serious Game (EQF2) **(WP2)**

2. "Algoritmisk Bias"-kursus (videregående uddannelser) – bLearning-tilgang: a) synkrone sessioner (helst face-to-face), b) asynkrone eLearning-sessioner i teoretiske komponenter, suppleret med et digitalt *serious game* (dvs. et avanceret læringsspil) som formativ evaluering og online summativ evaluering, tilgængelig som .scorm og klar til at blive uploadet i ethvert LMS til studerende, der er tilmeldt Big Data-, AI-, machine learning- og/eller deep learning-relaterede kurser. **(WP3)**

3. "Algoritmisk Bias toolkit til synkrone sessioner" – til lærere/professorer/lektorer, der implementerer de synkrone sessioner (face-to-face eller online), og der supplerer de asynkrone eLearning-sessioner. **(WP3)**
4. En guide til at øge evnen hos universitetsadministratorer/-ledelse af uddannelsesafdelinger med ansvar for digital uddannelse til at fremme digital uddannelse vedr. algoritmisk bias. **(WP3)**
5. Webinarer for at fremme *peer-learning* og diskutere de videregående uddannelsesinstitutioners rolle i tech-uddannelse med henblik på at fremme tværfaglighed inden for samfundsvidenskab i tech-uddannelser og for at diskutere tilgange til interoperabiliteten med udskoling samt ungdomsuddannelse og voksenundervisning. (**WP3)**
6. Selvstudium i "Etisk AI" (EQF4) for voksne elever, der følger undervisningen online og asynkront. **(WP4)**
7. Digitalt *Serious game* (EQF2) for unge mellem 12 og 18 år (primært udsatte unge/unge kvinder) for at fremme kønsrepræsentation i STEM allerede fra de yngre år. **(WP4)**
8. Værktøjskasse til at støtte voksne og unge i at opkvalificere sig til etisk AI. **(WP4)**
9. Politiske anbefalinger om anerkendelse af *micro-credentials* (dvs. kursusbeviser for enkeltstående, ofte kortere, digitale kurser) for voksne elever vedr. adgangen til videregående uddannelseskurser inden for teknologiske områder**. (WP4)**

1.4. CHARLIEs målgrupper

CHARLIE-projektets primære målgrupper er:
- Videregående uddannelsesinstitutioner, der udbyder læringsmuligheder eller kurser inden for Big Data, AI, machine learning, deep learning samt institutionernes administratorer/ledelser af deres forskellige uddannelsesafdelinger.
- Studerende på videregående uddannelser, der er tilmeldt kurser inden for Big Data, AI, machine learning eller deep learning.
- Undervisere/professorer/lektorer på videregående uddannelser fra samfundsvidenskab og/eller tech-uddannelser.
- Voksenundervisningsorganisationer/-institutioner og deres personale.
- Voksne i alle aldre og med forskellige socioøkonomiske baggrunde, der sigter mod at opnå højere kvalifikationsniveauer, som er relevante for arbejdsmarkedet og for aktiv deltagelse i samfundet.
- Ungdomsorganisationer/-institutioner og deres personale.
- Unge (12-18 år) - især unge kvinder og unge fra socialt udsatte miljøer (som oplever barrierer relateret til uddannelsessystemet, kulturelle forskelle, og/eller

sociale barrierer, økonomiske barrierer, diskrimination samt geografiske barrierer).

Hovedmålgrupper for kommunikations- og projektfremmende aktiviteter og resultater:
- Nationale lovgivere og politiske beslutningstagere;
- Den brede offentlighed
- Videregående uddannelsesinstitutioner (universitetslærere, studerende (bachelor- og kandidatstuderende) samt universitetsadministratorer)
- Udbydere af voksenuddannelse
- Ungdomsorganisationer/-institutioner
- Udbydere af erhvervsuddannelser
- AI-fora og -platforme
- NGO'er, der arbejder med fordomme og diskrimination
- Udbydere af uddannelsesteknologi
- Massemedie-virksomheder (tv, e-aviser, radiokanaler osv.)
- AI/ML SMV'er og større AI/ML-virksomheder
- Sammenslutninger af udbydere af videregående uddannelser, herunder relaterede samarbejdsorganisationer til fremme af best practice på EU-niveau og udbredelse til deres universitetsmedlemmer (f.eks. European University Association, Science Business Network, Triple Helix Association, AMBA, University-Industry Innovation Network, European Business Angels Network (til investering i etiske AI-innovationer)
- Studenterorganisationer i EU-området
- Sammenslutninger, som beskæftiger sig med monitorering/udvikling af kompetencer (f.eks. DigCompEdu-netværk).

1.5. Formålene med WP2

De specifikke formål for denne WP vedr. kompetencematricer for "Algoritmisk Bias" er:

- at skabe en fælles forståelse af formålet med og læringsmålene for et læringsforløb, der vil blive udviklet til studerende i kurset "Algoritmisk Bias" på videregående uddannelser;

- at skabe en fælles forståelse af formålet med og læringsmålene for et læringsforløbet "Etisk AI Micro-Credential", der vil blive udviklet til elever inden for voksenundervisning;

- at skabe en fælles forståelse af formålet med og læringsmålene for et læringsforløb, der vil blive udviklet til unge (12-18 år) i form af et "serious game" om etisk AI.

Disse formål bidrager til de generelle mål ved at sætte rammerne for de efterfølgende aktiviteter, så lærere, mentorer og elever klart kan se målene for de forskellige læringsforløb, og så højere læreanstalter, voksen- og ungdomsuddannelsesstederne bliver bedre i stand til at tilbyde læringsmuligheder, der opfylder samfundets behov, men som også er skræddersyet til deltagernes læringsbehov.

1.6. Forventede resultater af aktiviteterne i WP2 samt af deltagerne

1.6.1. Kompetencematrix EQF 6 for kurset "Algoritmisk Bias" (studerende på videregående uddannelser)

En tilgang til læringsudbytte, 2 ECVS-point, akkrediterings-anbefalinger, optagelseskrav, kontakttimer, samlet arbejdsbelastning, integration af EntreComp-kompetencer (f.eks. etisk og bæredygtig tænkning), DigComp 2.0-kompetencer (f.eks. beskyttelse af sundhed og velvære) og GrenComp-kompetencer (f.eks. understøttelse af retfærdighed).

Kompetenceenheder:
CU1 – Algoritme-modeller og begrænsninger UIB
CU2 – Data-fairness og bias i AI UA
CU3 – Privatlivets fred og bekvemmelighed i AI UA
CU4 – AI-etik, en praktisk tilgang VAMK
CU5 – Casestudier og projekter VAMK

1.6.2. Kompetencematrix EQF 4 for "Etisk AI Micro-Credential"

En tilgang til læringsudbytte, mulige ECVS-point, akkrediteringsanbefalinger, optagelseskrav, kontakttimer, samlet arbejdsbelastning, integration af EntreComp-kompetencer (f.eks. etisk og bæredygtig tænkning), DigComp 2.0-kompetencer (f.eks. beskyttelse af sundhed og velvære) og GrenComp-kompetencer (f.eks. understøttelse af retfærdighed).

Kompetenceenheder:
CU1 – Hvad er algoritmisk bias? HELIX
CU2 – Ikke-skadelighed HELIX
CU3 – Ansvarlighed ITC
CU4 – Gennemsigtighed ITC
CU5 – Menneskerettigheder og retfærdighed ITC
CU6 – AI-etik, en praktisk tilgang HELIX

1.6.3. Læringsudbytte for elever i udskoling og ungdomsuddannelse EQF 2 (serious game)

En tilgang til læringsudbytte, optagelseskrav, kontakttimer, integration af EntreComp-kompetencer (f.eks. etisk og bæredygtig tænkning), DigComp 2.0-kompetencer (f.eks. beskyttelse af sundhed og velvære) og GrenComp-kompetencer (f.eks. understøttelse af retfærdighed). Generelt indhold: Systemiske uligheder i samfundet, grundlæggende mekanismer i systemer med kunstig intelligens, politiske dagsordener, AI's indvirkning globalt.

2. INTRODUKTION TIL ALGORITMER, AI, BIAS OG ETIK

Algoritmer spiller en afgørende rolle i det moderne samfund, da de hjælper med forskellige services som f.eks. at anbefale sange, film eller venner (Paraschakis, 2018; Milano et al., 2020). De bruges også i skoler, på hospitaler, i finansielle institutioner, ved domstole og i regeringer til at træffe vigtige beslutninger (Obermeyer et al., 2019). Men brugen af algoritmer kan indebære etiske risici (Floridi et al., 2018).

Algoritmer kan være forudindtagede, altså *biased*, hvilket f.eks. kan påvirke oversættelser (Prates et al., 2020) eller jobannoncer (Lambrecht og Tucker, 2019). For eksempel kan job med højere lønninger blive vist mere til mænd end til kvinder. Det kan også føre til uretfærdig behandling i sundhedsvæsenet, hvor patienter med lys hud får bedre pleje end patienter med mørk hud (Obermeyer et al. 2019). Mens man forsøger at løse disse problemer, bliver der ved med at dukke nye op.

Siden 2012 har der været meget fokus på kunstig intelligens (AI) og dens etiske konsekvenser (Perrault et al. 2019). I 2016 forsøgte man i en undersøgelse at kortlægge disse problemer (Mittelstadt et al. 2016), men feltet er vokset og har ændret sig meget siden da. Regeringer, organisationer og virksomheder er blevet mere involveret i diskussionen om "fair" og "etisk" AI samt algoritmer (Wong 2019).

Bias i teknologi kan have skadelige virkninger, selv hvis det er utilsigtet, og kan udfordre offentlighedens tillid til AI. For at sikre, at AI er trygt og sikkert, skal vi overveje dens specifikke og potentielle indvirkning på samfundet. Den nuværende indsats for at håndtere AI-bias fokuserer på beregningsmæssige aspekter som datarepræsentativitet og retfærdighed i maskinlæringsalgoritmer. Men menneskelige, institutionelle og samfundsmæssige faktorer bidrager også til AI-bias og bliver ofte overset. For

at kunne tackle denne udfordring er vi nødt til at udvide vores perspektiv og undersøge, hvordan AI både skabes og påvirker samfundet.

Troværdig og ansvarlig AI handler ikke kun om, hvorvidt et AI-system er forudindtaget, retfærdigt eller etisk, men også om, hvorvidt det gør, hvad det påstår. En praktisk tilgang til gennemsigtighed, datasæt og test, evaluering, validering og verifikation (TEVV) er vigtig. Menneskelige faktorer, inddragende designteknikker, adgange for flere interessenter og *human-in-the-loop* hjælper også med at mindske risikoen for AI-bias. Men disse fremgangsmåder alene giver ikke en komplet løsning. Vi har brug for input og vejledning fra et bredere socioteknisk perspektiv, der forbinder disse tilgange med samfundsmæssige værdier.

Eksperter i troværdig og ansvarlig AI anbefaler at operationalisere disse værdier og skabe nye normer for AI-udvikling og -implementering. Dette dokument, sammen med det arbejde, der er udført af partnerne i CHARLIE-projektet om algoritme- og AI-bias, anlægger et socio-teknisk perspektiv.

Bias er ikke unikt for AI, og det er umuligt at opnå fuldstændig unbiased AI-systemer. CHARLIE-projektet har til hensigt at udvikle nogle ressourcer og metoder til at skabe større forbedringer i praksisserne for at identificere, forstå, måle, styre og reducere bias. For at nå dette mål har vi brug for fleksible teknikker, der kan anvendes på tværs af kontekster og kommunikeres til forskellige interessentgrupper.

3. KOMPETENCEMATRIX: EN TEORETISK TILGANG

En kompetencematrix, også kendt som en færdighedsmatrix eller kompetenceramme, er et værktøj, der bruges til at identificere, kortlægge og vurdere færdigheder, viden og evner hos enkeltpersoner i en organisation eller uddannelsesinstitution (Stewart og Bartrum, 1999). Den giver en visuel repræsentation af de kompetencer, der kræves til specifikke roller, projekter, opgaver eller akademiske uddannelser, og hjælper med at identificere de nuværende kompetenceniveauer hos medarbejdere, teammedlemmer, studerende eller andre involverede (Cottrell, 2018).

Kendetegn ved en kompetencematrix

En kompetencematrix er kendetegnet ved flere, vigtige punkter:

- Gitter- eller tabelformat: En kompetencematrix er typisk struktureret som et gitter eller en tabel med rækker, der repræsenterer enkeltpersoner, og kolonner, der repræsenterer de nødvendige kompetencer eller færdigheder (Mansfield, 2009).

- Vurderinger eller færdighedsniveauer: Hver celle i matricen indeholder en vurdering eller et færdighedsniveau for den tilsvarende kompetence og person. Bedømmelsessystemet kan være numerisk, farvekodet eller baseret på beskrivende termer som "begynder", "mellemniveau" eller "ekspert" (Cottrell, 2018).

- Kan skræddersys og tilpasses: Kompetencematricer kan tilpasses, så de passer til en organisations eller uddannelsesinstitutions specifikke behov, og de kan nemt opdateres, så de afspejler ændringer i roller, ansvarsområder eller kompetencekrav (Stewart og Bartrum, 1999).

- Omfattende: En kompetencematrix bør dække alle de relevante kompetencer, der kræves til en bestemt rolle, et projekt eller en akademisk uddannelse, og bør omfatte både tekniske og mere 'bløde' færdigheder (Mansfield, 2009).

En kompetencematrix' hovedformål

En kompetencematrix tjener flere vigtige formål i en organisation eller uddannelsesinstitution:

- Identificering af kvalifikationshuller: En kompetencematrix kan, ved at give et klart overblik over de færdigheder og kompetencer, der kræves til en bestemt rolle eller et projekt, hjælpe med at identificere områder, hvor enkeltpersoner kan have brug for yderligere udvikling eller undervisning (Stewart og Bartrum, 1999).

- Støtte til medarbejderes eller studerendes udvikling: En kompetencematrix kan bruges til at sætte retningen for planer om personlig udvikling og præge uddannelses- og udviklingsinitiativer ved at hjælpe enkeltpersoner med at identificere deres styrker og svagheder og sætte mål for deres fremtidige udvikling (Cottrell, 2018).

- Optimering af ressourceallokering: En kompetencematrix kan hjælpe organisationer og uddannelsesinstitutioner med at allokere ressourcer mere effektivt ved at identificere de mest egnede personer til specifikke roller eller projekter baseret på deres færdigheder og kompetencer (Mansfield, 2009).

- Forbedring af præstationer: Ved at hjælpe med at sikre, at enkeltpersoner har de rette færdigheder og kompetencer til deres roller, kan en kompetencematrix bidrage til at forbedre den samlede præstation i en organisation eller uddannelsesinstitution (Stewart og Bartrum, 1999).

- Fremme af kommunikation og samarbejde: En kompetencematrix kan fungere som et fælles sprog til at diskutere færdigheder og kompetencer inden for en organisation eller uddannelsesinstitution, hvilket fremmer større forståelse og samarbejde blandt teammedlemmer, afdelinger eller akademiske afdelinger (Cottrell, 2018).

Anvendelse af en kompetencematrix i forskellige kontekster

Kompetencematricer er blevet anvendt i forskellige sammenhænge, som f.eks. i erhvervslivet, i det offentlige og på videregående uddannelser, for at opfylde en række formål:

- Arbejdsplanlægning: Kompetencematricer kan bruges til at understøtte den strategiske planlægning af de ansattes tid ved at identificere de færdigheder og kompetencer, der er nødvendige for at nå organisatoriske mål og målsætninger, og sikre, at disse afspejles i rekrutterings-, undervisnings- og udviklingsinitiativer (Stewart og Bartrum, 1999).

- Performance management: Kompetencematricer kan integreres i performance management-systemer, der giver en ramme for at opstille forventninger til præstationer, gennemføre præstationsevalueringer og identificere områder, der kan forbedres (Mansfield, 2009).

- Planlægning af succession: Kompetencematricer kan hjælpe i arbejdet med successionsplanlægning ved at identificere de

færdigheder og kompetencer, der kræves til lederroller, og hjælpe med at identificere og udvikle potentielle efterfølgere i en organisation eller uddannelsesinstitution (Cottrell, 2018).

- Planlægning og udvikling af uddannelsesforløb: På højere læreanstalter kan kompetencematricer bruges til at styre planlægning og udvikling af uddannelsesforløb og sikre, at akademiske uddannelser er designet til at udvikle de færdigheder og kompetencer, der kræves af studerende for at få succes inden for deres valgte områder (Billett, 2009). Denne anvendelse, og de følgende tre, er omdrejningspunkterne for dette dokument.

- Udvikling af fakultet og personale: Kompetencematricer kan anvendes til at støtte den faglige udvikling af undervisere og personale på videregående uddannelsesinstitutioner ved at identificere de færdigheder og kompetencer, der kræves for effektiv undervisning, forskning og administrative roller, og præge udformningen af uddannelses- og udviklingsprogrammer (Eraut, 2004).

- Dannelse af forskerteams: Kompetencematricer kan bruges til at gøre det nemmere at danne forskningsteams på videregående uddannelsesinstitutioner ved at identificere personer med komplementære færdigheder og kompetencer og fremme tværfagligt samarbejde (Börner et al., 2010).

- Rådgivning og støtte til studerende: Kompetencematricer kan bruges til at fremme studievejledning og andre understøttende services på videregående uddannelsesinstitutioner ved at hjælpe rådgivere med at identificere studerendes styrker og svagheder og guide dem mod passende ressourcer, interventioner eller akademiske uddannelsesveje (Cuseo, 2010).

Kompetencematricens udfordringer og begrænsninger

Selvom kompetencematricer kan give værdifuld viden og understøtte en række strategiske mål, har de også nogle udfordringer og begrænsninger:

- Subjektivitet: Kompetencevurderinger kan være subjektive og kan påvirkes af faktorer som personlige fordomme, kulturelle forskelle eller forskellige fortolkninger af kompetencedefinitioner (Stewart og Bartrum, 1999).

- Tids- og ressourcekrævende: At udvikle og vedligeholde en omfattende kompetencematrix kan være tids- og ressourcekrævende,

især i store organisationer eller uddannelsesinstitutioner med forskellige roller og ansvarsområder (Mansfield, 2009).

- Overdreven opmærksomhed på tekniske færdigheder: Kompetencematricer kan lægge større vægt på tekniske færdigheder og kompetencer på bekostning af mere 'bløde' færdigheder, såsom kommunikation, teamwork eller problemløsning, som dog ofte er afgørende for et godt resultat i mange funktioner og sammenhænge (Cottrell, 2018).

- Rigiditet: Kompetencematricer kan opfattes som rigide og ufleksible, hvilket kan begrænse kreativitet, innovation eller fleksibilitet i en organisation eller uddannelsesinstitution (Eraut, 2004).

På trods af disse udfordringer og begrænsninger kan kompetencematricer være et værdifuldt værktøj for organisationer og uddannelsesinstitutioner, der ønsker at optimere ressourceallokeringen, støtte den individuelle udvikling og forbedre den samlede præstation. Ved at anvende en systematisk og omfattende tilgang til at identificere, kortlægge og vurdere færdigheder og kompetencer og inddrage en række perspektiver og interessenter i udviklingen og vedligeholdelsen af matricen kan organisationer og uddannelsesinstitutioner maksimere fordelene ved denne tilgang og samtidig minimere de potentielle ulemper (Billett, 2009).

4. CHARLIE WP2-MATRICER

Kompetencematricen i EQF6-kurset "Algoritmisk Bias" fungerer som grundlag for to andre vigtige dele: EQF4-kompetencematricen i kurset "Etisk AI Micro-Credential" og "Læringsudbytte for voksne og unge EQF2 (Serious Game)". Disse tre dele er indbyrdes forbundne og designet til at supplere og underbygge hinanden, hvilket sikrer en dyb forståelse og omfattende anvendelse af etiske AI-principper på tværs af forskellige uddannelsesniveauer og kontekster.

Kompetencematricen for kurset "Algoritmisk Bias" danner grundlaget for udviklingen af EQF4-kompetencematricen for "Etisk AI Micro-Credential". Idet førstnævnte kursus udgør et stærkt fundament i forståelsen af algoritmisk bias og konsekvenserne heraf, giver det deltagerne mulighed for at dykke dybere ned i de etiske dimensioner af AI, når de går videre til EQF4-niveau. EQF4-kompetencematricen for "Etisk AI Micro-Credential" bygger på de først opnåede kompetencer og udvides derefter til at omfatte bredere, etiske

overvejelser og praktisk anvendelse af etiske AI-principper i forskellige sektorer og scenarier.

Samtidig præger kompetencematricen for kurset "Algoritmisk Bias" også "Læringsudbytte for voksne og unge EQF2 (Serious Game)". Dette sikrer, at de grundlæggende koncepter og færdigheder, der udvikles i kurset, er til at forstå samt engagerende for elever på EQF2-niveau. Serious game-formatet er designet til at introducere vigtige emner på en måde, der både er underholdende og lærerig, hvilket fremmer en dybere forståelse af algoritmisk bias og etiske AI-principper blandt en bredere modtagergruppe.

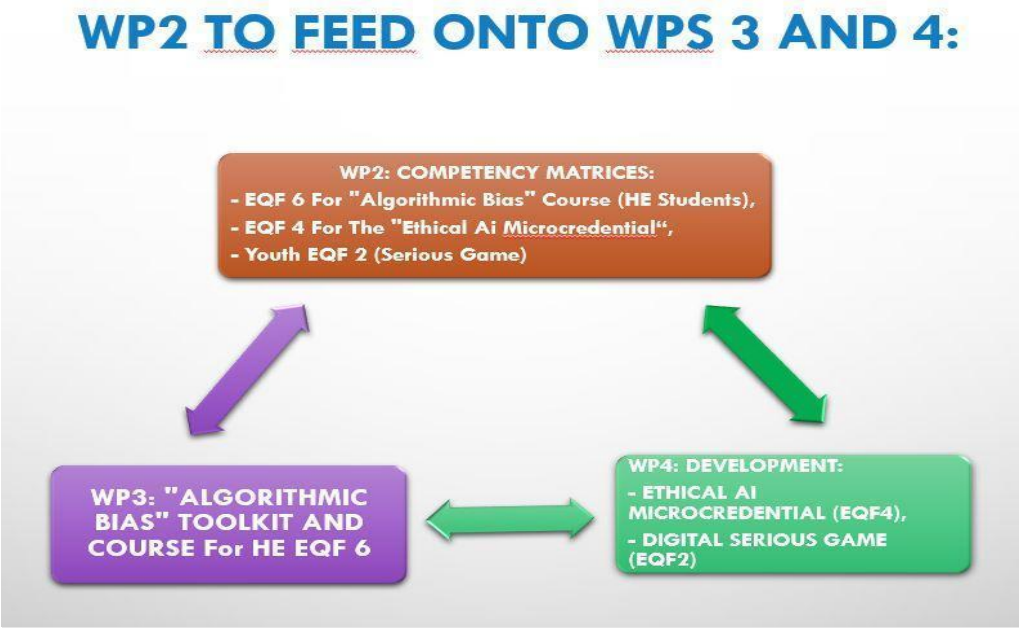


Illustration 1. Indbyrdes forbindelser mellem projektets WP'er og resultater

Afslutningsvis kan man konkludere, at de tre elementer - kompetencematricen for kurset "Algoritmisk Bias", EQF4-kompetencematricen for "Etisk AI Micro-Credential" og "Læringsudbytte for voksne og unge EQF2 (Serious Game)" - er tæt forbundne, hvilket også fremgår af illustration 1. Denne forbundne struktur gør det muligt at anlægge en bred tilgang på flere niveauer til at lære om algoritmisk bias og etisk AI, hvilket sikrer, at elever på tværs af forskellige uddannelsesbaggrunde og aldersgrupper kan udvikle den viden og de færdigheder, der er nødvendige for at navigere i udfordringerne og mulighederne ved AI på en etisk og ansvarlig måde. Resultaterne af WP2 vil derudover også indgå i WP3 og WP4, hvilket yderligere illustrerer det tætte forhold mellem disse elementer, som vist i illustration 1.

Denne fine integration af WP2, WP3 og WP4 resulterer i en solid og sammenhængende uddannelsesstruktur, hvor de meget vigtige aspekter af algoritmisk bias og etisk AI behandles, hvilket fremmer en god forståelse og anvendelse af disse principper på tværs af forskellige kontekster og deltagerprofiler.

4.1. KURSET "ALGORITMISK BIAS" EQF6

Dette uddannelsesforløb er designet til at give en grundig og omfattende forståelse af algoritmiske bias og deres implikationer i nutidens datadrevne samfund. Da algoritmer spiller en væsentlig rolle i vores liv og præger beslutningstagning på tværs af forskellige sektorer, som f.eks. sundhedsvæsnet, uddannelse, finanssektoren og regeringer, er det afgørende for fagfolk og akademikere/studerende at udvikle en solid forståelse af potentielle bias i diverse systemer.

Formålet med dette kursus er at give deltagerne den viden og de færdigheder, der er nødvendige for at identificere, afbøde effekterne af og håndtere algoritmiske bias. Gennem en kombination af teoretisk og grundlæggende viden samt forskellige former for praktisk anvendelse vil deltagerne opnå indsigt i de etiske, sociale og tekniske aspekter af algoritmisk bias.

Disse formål bidrager til de generelle mål ved at sætte rammerne for de følgende aktiviteter, så lærere, mentorer og kursusdeltagere kan få en klar forestilling om målene for de forskellige uddannelsesveje, og så mulighederne øges hos videregående læreanstalter samt voksen- og ungdomsuddannelsessteder for at tilbyde læringsmuligheder, der opfylder samfundets behov, men som også er skræddersyet til kursusdeltagernes læringsbehov. Helt konkret er de forventede resultater af denne WP rettet mod at:

1) øge de videregående uddannelsesinstitutioners mulighed at tilbyde deres studerende online kurser, der opfylder samfundets behov, men som også er skræddersyet til de studerendes læringsbehov – og hvor de videregående uddannelsesinstitutioner drager fordel af tankegangen om læringsudbytte, hvorved undervisere og studerende har et fælles grundlag for de forventede resultater ved afslutningen af et læringsforløb;

2) øge tech-studerendes sociale og etiske kompetencer, så de kan håndtere AI/ML-teknologi på en positiv, kritisk og etisk måde, og så universitetsstuderende drager fordel af tankegangen om læringsudbytte, hvor de klart ved, hvad der forventes af dem ved afslutningen af læringsforløbet;

3) udstyre undervisere/professorer med digitale og engagerende tilgange til effektiv undervisning i emnet (især i onlineundervisning) - da den elevcentrerede tilgang er udgangspunktet for designet af læringsmetoderne;
4) skabe synergier mellem videregående uddannelsesinstitutioner og voksen- samt ungdomsuddannelserne på regionalt/lokalt niveau inden for uddannelse i AI-etik ved at fastsætte læringsudbytte for komplementære OER (*Open Educational Ressources*, dvs. gratis og frit tilgængelige online-ressourcer), der er målrettet forskellige niveauer (EQF 6, 4, 2) inden for det samme vidensområde;
5) styrke overførbarheden af akademiske kurser om AI-bias til voksen- samt ungdomsuddannelserne ved at etablere nogle forbindelser mellem de forskellige læringsresultater, som kan danne udgangspunkt for et fælles grundlag for diskussion af læringsforløbenes potentiale for at kunne overføres;
6) øge bevidstheden om emnet på samfundsniveau, da det er det første skridt til at få målrettede OER'er ud til forskellige samfundslag og målgrupper.

Efter endt kursus vil deltagerne have udviklet en omfattende forståelse af algoritmiske bias, deres mulige indvirkning på forskellige sektorer og de strategier og værktøjer, der er til rådighed for at håndtere disse bias. Denne viden vil være uvurderlig for fagfolk og akademikere/studerende, der arbejder inden for områder, hvor algoritmer bruges til beslutningstagning, og vil hjælpe med at sikre mere lige og retfærdige resultater i en datadrevet verden.

4.1.1. Indhold i kurset "Algoritmisk Bias"

Kurset er struktureret omkring fem primære kompetenceenheder (*Competence Units*, her forkortet CU'er), der hver især er designet til at give deltagerne den viden og de færdigheder, der er nødvendige for at navigere i udfordringerne og mulighederne inden for algoritmisk bias.

CU1 – Algoritmemodeller og begrænsninger: Denne del introducerer de grundlæggende begreber inden for algoritmer, deres modeller og begrænsninger. Deltagerne vil få en forståelse af den rolle, algoritmer spiller i forskellige sektorer, og de potentielle faldgruber, der opstår ved brugen af dem. Emnerne vil omfatte algoritmisk kompleksitet, optimering og begrænsningerne ved algoritmisk beslutningstagning.

CU2 – Data-fairness og bias i AI: Denne del vil dykke ned i begreberne *fairness* og *bias* i AI-systemer og udforske de forskellige kilder til og udtryk af bias i algoritmiske processer. Deltagerne vil lære om metoder til at identificere og måle bias i data, og de vil få strategier til at håndtere og afbøde effekterne af disse bias for at sikre mere retfærdige resultater.

CU3 – Privatlivets fred og bekvemmelighed vedr. AI: I denne del vil deltagerne få mulighed for at undersøge kompromiserne mellem opretholdelse af privatlivets fred og bekvemmelighed i AI-applikationer. Der vil blive redegjort for udfordringerne ved at opretholde privatlivets fred hos brugerne, samtidig med at der bliver leveret personaliserede og effektive services. Der vil endvidere blive undervist i eksisterende og nye teknologier, der er designet til at bringe disse modstridende interesser i balance.

CU4 – AI-etik, en praktisk tilgang: Denne del vil fokusere på den praktiske anvendelse af etiske principper i AI-udvikling og -implementering. Deltagerne vil undersøge forskellige etiske rammer og retningslinjer og lære, hvordan man anvender i virkeligheden, hvor AI-systemer indgår. Deltagerne vil også få viden om vigtigheden af gennemsigtighed, ansvarlighed og interessenters involvering i etisk AI-udvikling.

CU5 – Casestudier og projekter: Den sidste del giver deltagerne mulighed for at anvende deres viden og færdigheder i casestudier samt et *hands-on*-projekt. Gennem denne praktiske tilgang vil deltagerne få en dybere forståelse af kompleksiteten og nuancerne i algoritmisk bias samt de strategier og værktøjer, der er til rådighed for at løse disse problemstillinger.

4.1.2. Beskrivelse af målgruppe og EQF6-niveau

Den europæiske referenceramme for kvalifikationer (EQF) er en fælles referenceramme, der skaber forbindelse mellem kvalifikationssystemerne i de forskellige europæiske lande og gør det lettere for enkeltpersoner at få deres kvalifikationer anerkendt i hele Europa (Europa-Kommissionen, 2023). EQF består af otte referenceniveauer, som hver især er defineret af nogle deskriptorer, der angiver de læringsresultater, færdigheder, kompetencer og den selvstændighed, der forventes på det pågældende niveau. EQF-niveau 6 (EQF6) svarer til kvalifikationer opnået på niveau med en bachelorgrad eller en højere erhvervskvalifikation i det europæiske område for videregående

uddannelse (EHEA). EQF6-niveauet kan beskrives på følgende måde (European Commission, 2023; CEDEFOP, 2023):

- Avanceret viden: EQF6-kvalifikationer kræver, at man har avanceret viden om et specifikt område, hvor man kan vise, at man har en kritisk forståelse af teorier, principper og metoder. Dette vidensniveau er nødvendigt for at udvikle nye idéer og løse komplekse problemer.

- Kognitive færdigheder: På EQF6 skal man udvikle evnen til at anvende sin viden i nye eller ukendte situationer, analysere komplekse problemer og sammenfatte information fra forskellige kilder. Dette indebærer evnen til kritisk tænkning, problemløsning, beslutningstagning og kreativ tænkning.

- Praktiske færdigheder: EQF6-kvalifikationer omfatter udviklingen af avancerede praktiske færdigheder, der gør det muligt at styre komplekse tekniske eller faglige aktiviteter, projekter eller processer. Disse færdigheder kan omfatte forskning, projektledelse eller evnen til at bruge specialiserede værktøjer og teknikker effektivt.

- Selvstændighed og ansvar: Elever og studerende på EQF6-niveau forventes at udvise en høj grad af selvstændighed og ansvarlighed i deres arbejde. Det indebærer, at de tager ansvar for egen læring og faglige udvikling, leder og superviserer andres arbejde og træffer kvalificerede beslutninger baseret på etiske og sociale overvejelser.

- Kommunikation og samarbejde: EQF6-kvalifikationer kræver, at man udvikler effektive kommunikations- og samarbejdsevner, så man kan præsentere klare og detaljerede oplysninger, argumenter eller forslag til både specialister og ikke-specialister. Dette omfatter skriftlige, mundtlige og interpersonelle kommunikationsevner samt evnen til at arbejde effektivt i teams og på tværs af discipliner.

- Livslang læring: På EQF6 skal man udvikle evnen til at involvere sig i livslang læring, reflektere over egen læring og opdage områder for yderligere udvikling. Dette omfatter evnen til at tilpasse sig nye situationer, teknologier eller faglige miljøer og til at involvere sig i fortsat faglig udvikling.

I et forsøg på at overføre EQF6's karakteristika til emnet for CHARLIE-projektet er her nogle eksempler på kompetencer inden for algoritmebias, der svarer til EQF6-niveauet:

- Avanceret viden: Deltagerne skal have en dyb forståelse af algoritmisk bias, dens kilder og dens indvirkning på AI/ML-systemer. De skal være fortrolige med forskellige typer bias, retfærdighedsmålinger og de etiske implikationer af *biased* algoritmer.

- Kognitive færdigheder: Deltagerne skal være i stand til at identificere og analysere tilfælde af algoritmisk bias i virkelige situationer, kritisk vurdere AI/ML-systemers retfærdighed og foreslå metoder til at mindske eller eliminere bias.

- Praktiske færdigheder: Deltagerne skal udvikle evnen til at anvende teknikker til at opdage og mindske algoritmisk bias; det kunne være teknikker såsom *re-sampling*, *re-weighting* eller *adversarial training*. De bør også blive i stand til at evaluere effektiviteten af disse teknikker til at reducere bias og øge retfærdighed.

- Selvstændighed og ansvar: Deltagerne skal være i stand til at træffe kvalificerede beslutninger om etisk brug af AI/ML-systemer under hensyntagen til de potentielle risici og fordele, der er forbundet med algoritmisk bias. De skal også være i stand til at tage ansvar for konsekvenserne af deres beslutninger for enkeltpersoner, lokalsamfund og samfundet som helhed.

- Kommunikation og samarbejde: Deltagerne skal være i stand til at kommunikere deres forståelse af algoritmisk bias og dens konsekvenser til målgrupper med teknisk og ikke-teknisk baggrund. De skal også være i stand til at samarbejde effektivt med tværfaglige teams om at løse problemer om retfærdighed, gennemsigtighed og ansvarlighed i AI/ML-systemer.

- Livslang læring: Deltagerne skal bestræbe sig på at holde sig ajour med den nyeste forskning, metoder og værktøjer relateret til algoritmisk bias, retfærdighed og etik i AI/ML. De skal være i stand til at tilpasse deres viden og færdigheder til nye udfordringer og muligheder inden for området.

Samlet set er EQF6-kvalifikationer kendetegnet ved avanceret viden, kognitive og praktiske færdigheder, selvstændighed og ansvar, effektiv kommunikation og samarbejde samt en forpligtelse til livslang læring. Disse karakteristika og funktioner sikrer, at dimittender på dette niveau er godt rustet til at få succes i deres valgte erhverv eller fortsætte deres uddannelse på højere niveauer, såsom at tage en kandidatgrad (EQF7) eller en ph.d.-grad (EQF8).

4.1.3. Generel kompetencematrix for kurset "Algoritmisk Bias" EQF6

LÆRINGSUDBYTTE		
	Efter endt læringsforløb vil deltageren være i stand til at	
VIDEN	**FÆRDIGHEDER**	**HOLDNINGER**
1. Forstå hvordan AI's tværfaglige karakter og algoritmernes rolle præger udformningen af forskellige forhold i samfund, økonomi og teknologi. 2. Opnå en omfattende forståelse af de etiske, juridiske og sociale implikationer af AI-systemer, herunder spørgsmål relateret til retfærdighed, privatlivets fred og bekvemmelighed. 3. Udvikle evnen til kritisk tænkning og problemløsning med henblik på at identificere, analysere og håndtere algoritmisk bias og andre etiske udfordringer i AI. 4. Evaluere effektiviteten af forskellige tilgange til at afbøde effekterne af algoritmisk bias og fremme retfærdighed i AI-systemer under hensyn til begrænsninger og kompromiser. 5. Forstå vigtigheden af en brugercentreret tilgang til design og implementering af AI-systemer med respekt for privatlivets fred, autonomi og individuelle præferencer. 6. Anerkende betydningen af tværfagligt samarbejde og inddragelse af interessenter i udviklingen og implementeringen af etiske AI-løsninger. 7. Udvikle et stærkt fundament for de etiske teorier og rammer, der gælder for AI, og lære at anvende disse principper i den virkelige verden. 8. Opnå en forståelse for vigtigheden af gennemsigtighed, forklarbarhed og ansvarlighed i udviklingen og implementeringen af AI-systemer. 9. Forstå den rolle, som forordninger, branchestandarder og best practice spiller i udformningen af etisk AI-udvikling og -implementering.	1. Analysere og evaluere AI-algoritmer og -modeller for at identificere potentielle bias og begrænsninger og foreslå forbedringer eller alternativer, når det er nødvendigt. 2. Anvende forskellige metoder og teknikker til at identificere, måle og afbøde effekterne af bias i data og AI-systemer for at sikre retfærdighed og retfærdige resultater. 3. Bringe balance mellem opretholdelse af privatlivets fred og bekvemmelighed i AI-applikationer og træffe informerede beslutninger med hensyn til at beskytte brugernes privatliv og samtidig levere personaliserede og effektive services. 4. Implementere etiske principper, rammer og retningslinjer i design, udvikling og implementering af AI-systemer, som sikrer overholdelse af standarder for gennemsigtighed, ansvarlighed og inddragelse af interessenter. 5. Kritisk vurdere casestudier og scenarier fra den virkelige verden, der involverer AI-systemer, identificere etiske udfordringer og foreslå løsninger baseret på den viden og de færdigheder, der er erhvervet i kurset. 6. Samarbejde effektivt med tværfaglige teams og derved demonstrere en forståelse af de forskellige perspektiver og ekspertiser, der kræves for at håndtere de komplekse etiske udfordringer inden for AI. 7. Anvende brugercentrerede designprincipper og -strategier til at udvikle AI-systemer, der respekterer privatlivets fred, autonomi og individuelle præferencer.	1. Selvstændigt genkende og håndtere etiske udfordringer og bias i AI-systemer og tage ansvar for at sikre retfærdighed, privatlivets fred og bekvemmelighed i udviklingen og implementeringen af sådanne systemer. 2. Udvise en forpligtelse til etisk beslutningstagning i AI-relaterede projekter, anerkende de potentielle konsekvenser af deres handlinger og tage ansvar for deres valg. 3. Tage initiativ til at holde sig informeret om aktuelle og nye AI-etiske emner, regler og best practice og aktivt opsøge muligheder for kontinuerlig læring og faglig udvikling. 4. Samarbejde effektivt med tværfaglige teams, idet de bidrager med viden og ekspertise og samtidig respekterer og værdsætter andres perspektiver samt tager del i ansvaret for projekters resultater. 5. Udvise en ansvarsfølelse over for samfundet ved at udvikle og fremme AI-systemer, der imødekommer samfundsmæssige behov, minimerer skadelige virkninger og maksimerer fordele for forskellige interessenter. 6. Gå ind for gennemsigtighed, ansvarlighed og inddragelse af interessenter i udvikling og implementering af AI, tage ansvar for etiske overvejelser og opretholde en åben kommunikation med relevante parter. 7. Anvende etiske principper og retningslinjer konsekvent på tværs af forskellige AI-projekter og kontekster, demonstrere personligt engagement i etisk AI-udvikling og tage ansvar for at opretholde disse standarder. 8. Handle selvstændigt ved at identificere områder for forbedring af eller innovation i AI-systemer og tage initiativ til at foreslå og implementere løsninger, der adresserer etiske udfordringer og fremmer retfærdighed.

10. Udvikle effektive kommunikationsevner til at deltage i kvalificerede diskussioner og debatter om AI-etik, algoritmisk bias og relaterede emner med forskellige målgrupper.	8. Formidle kompleks AI-etik og algoritmiske bias-koncepter og løsninger klart og effektivt til forskellige målgrupper, både tekniske og ikke-tekniske. 9. Holde sig informeret om gældende og nye forordninger, branchestandarder og best practice i forbindelse med AI-etik, og anvende denne viden til at udvikle AI-systemer, der overholder reglerne. 10. Fokusere på kontinuerlig læring og refleksion og forsøge at forbedre og udvide deres forståelse af AI-etik, algoritmisk bias og andre relaterede emner i det hurtigt udviklende AI-landskab.	9. Erkende og håndtere personlige begrænsninger i viden og færdigheder relateret til AI-etik og tage ansvar for at søge støtte, feedback og muligheder for læring - alt efter behov. 10. Fremme en kultur med etisk bevidsthed og ansvar inden for deres professionelle og akademiske fællesskaber ved at dele viden, erfaringer og best practice relateret til AI-etik og algoritmisk bias.

4.1.4. Kompetencematrix for hver kompetenceenhed i "Algoritmisk Bias"-kurset EQF6

CU1 – Algoritme-modeller og begrænsninger EQF6 **LÆRINGSUDBYTTE**		
	Efter endt læringsforløb vil deltageren være i stand til at	
VIDEN	**FÆRDIGHEDER**	**HOLDNINGER**
Teoretisk/faglig viden i: CUK1.1. Forstå de grundlæggende begreber vedrørende algoritmer, deres modeller og begrænsninger. CUK1.2. Forstå algoritmers rolle i forskellige sektorer og de potentielle faldgruber, der opstår ved brugen af dem. CUK1.3. Forstå algoritmisk kompleksitet, optimering og begrænsninger ved algoritmisk beslutningstagning.	Færdighedsmål 1: Analysere og evaluere algoritmiske modeller 1.1. Vurdere styrker og svagheder ved forskellige algoritmiske modeller og deres anvendelighed i forskellige sammenhænge og problemer. 1.2. Kritisk evaluere algoritmers begrænsninger i beslutningsprocesser under hensyntagen til faktorer som beregningskompleksitet, datakvalitet og retfærdighed. Færdighedsmål 2: Anvende algoritmiske løsninger på problemer i den virkelige verden 2.1. Udvælge passende algoritmer til specifikke opgaver under hensyntagen til deres egnethed til det givne problem og de potentielle faldgruber forbundet med brugen af dem. 2.2. Implementere og finjustere algoritmer for at optimere deres funktion under hensyntagen til faktorer som effektivitet, nøjagtighed og skalerbarhed. Færdighedsmål 3: Afbøde effekterne af algoritmiske risici og bias 3.1. Identificere potentielle kilder til bias, fejl eller utilsigtede konsekvenser i algoritmiske beslutningsprocesser. 3.2. Udvikle og anvende strategier til at minimere algoritmiske risici og bias og dermed sikre mere præcise, retfærdige og etiske resultater i forskellige sektorer.	Holdningsmål 1: Etisk bevidsthed i algoritmeudvikling og -implementering 1.1. Udvise stor ansvarsfølelse i udviklingen og implementeringen af algoritmer under hensyntagen til den potentielle indvirkning på enkeltpersoner, lokalsamfund og samfundet som helhed. 1.2. Udvise forpligtelse til retfærdighed, gennemsigtighed og ansvarlighed i design og brug af algoritmer og søge at minimere potentielle bias og utilsigtede konsekvenser. Holdningsmål 2: Kritisk tænkning og refleksiv praksis 2.1. Have en kritisk tankegang, når man beskæftiger sig med algoritmer, deres modeller og begrænsninger, og konstant sætte spørgsmålstegn ved antagelser samt overveje alternative perspektiver. 2.2. Udøve en refleksiv praksis og evaluere personlige og professionelle erfaringer med algoritmer for at identificere områder, der kan forbedres, og løbende tage ved lære af det. Holdningsmål 3: Samarbejde og tværfaglig tilgang 3.1. Anerkende værdien af samarbejde og tværfaglige tilgange, når man tackler komplekse problemer fra den virkelige verden, hvor algoritmer er involvereret. 3.2. Opsøge muligheder for at arbejde med forskellige teams og lære af eksperter inden for forskellige områder for at udvikle gode, innovative løsninger, der adresserer algoritmiske udfordringer holistisk.

KOMPETENCE/ER INDEN FOR RAMMERNE AF DIGCOMP2.2.[1], ENTRECOMP[2] OG GREENCOMP[3], SOM CU1 ER FORBUNDET MED	FÆRDIGHEDER INDEN FOR RAMMERNE AF DIGCOMP2.2., ENTRECOMP OG GREENCOMP, SOM CU1 ER FORBUNDET MED	HOLDNINGER INDEN FOR RAMMERNE AF DIGCOMP2.2., ENTRECOMP OG GREENCOMP, SOM CU1 ER FORBUNDET MED
- Ved, at AI i sig selv hverken er godt eller dårligt. Det, der afgør, om resultaterne af et AI-system er positive eller negative for samfundet, er, hvordan AI-systemet er designet og brugt, af hvem og til hvilke formål (DIGCOMP2.2.). - Er bevidst om, at søgemaskiner, sociale medier og indholdsplatforme ofte bruger AI-algoritmer til at generere svar, der er tilpasset den enkelte bruger (f.eks. bliver brugerne ved med at få lignende resultater eller indhold). Dette omtales ofte som "personalisering" (DIGCOMP2.2.). - Er bevidst om, at AI-algoritmer fungerer på måder, der normalt ikke er synlige eller lette at forstå for brugerne. Dette kaldes ofte "black box"-beslutningstagning, da det kan være umuligt at spore, hvordan og hvorfor en algoritme kommer med specifikke forslag eller forudsigelser (DIGCOMP2.2.). - Ved, at alle EU-borgere har ret til ikke at være underlagt fuldautomatisk beslutningstagning (f.eks. hvis et automatisk system afviser en kreditansøgning, har kunden ret til at bede om, at beslutningen bliver gennemgået af en person). (DIGCOMP2.2.) - Er i stand til at identificere nogle eksempler på AI-systemer: produktanbefalinger (f.eks. på online shopping sites), stemmegenkendelse (f.eks. af virtuelle assistenter), billedgenkendelse (f.eks. til at opdage tumorer på røntgenbilleder) og	- Kan identificere områder, hvor AI kan være en fordel i hverdagen. I sundhedssektoren kan AI f.eks. bidrage til tidlig diagnosticering, mens man i landbruget kan bruge AI til at opdage skadedyrsangreb (DIGCOMP2.2.). - Ved, hvordan man formulerer søgeforespørgsler for at opnå det ønskede output, når man interagerer med samtaleagenter eller chatbots (f.eks. Siri, Alexa, Cortana, Google Assistant), f.eks. ved at vide, at forespørgslen skal være entydig og formuleres klart, hvis systemet skal kunne svare på det, som systemet bliver bedt om (DIGCOMP2.2).	- Være villig til at overveje etiske spørgsmål i forbindelse med AI-systemer (f.eks. i hvilke sammenhænge, såsom domfældelse af kriminelle, bør AI's anbefalinger ikke bruges uden menneskelig indgriben)? (DIGCOMP2.2.) - Afveje fordele og ulemper ved at bruge AI-drevne søgemaskiner (selvom de f.eks. kan hjælpe brugere med at finde den ønskede information, kan de kompromittere privatlivets fred og personlige data eller underkaste brugeren kommercielle interesser) (DIGCOMP2.2.). - Er indstillet på at blive ved med at lære, uddanne sig og holde sig ajour om AI (f.eks. at forstå, hvordan AI-algoritmer fungerer; at forstå, hvordan automatisk beslutningstagning kan være partisk; at skelne mellem realistisk og urealistisk AI; og at forstå forskellen mellem snæver kunstig intelligens, dvs. nutidens AI, der er i stand til at udføre simple opgaver som at spille spil, og generel kunstig intelligens, dvs. AI, der overgår menneskelig intelligens, hvilket stadig er science fiction) (DIGCOMP2.2).

[1] https://publications.jrc.ec.europa.eu/repository/handle/JRC128415
[2] https://joint-research-centre.ec.europa.eu/entrecomp-entrepreneurship-competence-framework/competence-areas-and-learning-progress_en
[3] https://joint-research-centre.ec.europa.eu/greencomp-european-sustainability-competence-framework_en

ansigtsgenkendelse (f.eks. i overvågningssystemer). (DIGCOMP2.2.) - Er bevidst om, at AI er et felt i konstant udvikling, og hvis fortsatte udvikling og indvirkning stadig er meget uklar (DIGCOMP2.2.)		

CU1 BESKRIVELSE AF VIDENSUDBYTTE

1.1. Forstå de grundlæggende begreber vedrørende algoritmer, deres modeller og begrænsninger. At forstå de grundlæggende begreber vedrørende algoritmer, deres modeller og begrænsninger indebærer en dyb fordybelse i de "grundlæggende byggesten i algoritmer, som er forankret i matematiske og beregningsmæssige teorier" (Smith, 2018, s. 45, frit oversat). Dette omfatter en forståelse af det iboende design, funktionaliteten og de potentielle begrænsninger, der påvirker ydeevne og nøjagtighed.

Algoritmer: Ifølge Knuth (1997, frit oversat) er "en algoritme et begrænset sæt regler, der udløser en sekvens af operationer til løsning af en bestemt type problem". Inden for datalogi og kunstig intelligens er disse procedurer grundlæggende for databehandling, beslutningstagning og automatisering.

Et eksempel: Den euklidiske algoritme er en banebrydende teknik til at finde den største fælles divisor (*greatest common divisor* - GCD) for to tal, og den demonstrerer princippet om proceduremæssig reduktion for at finde en løsning iterativt (Johnson, 2003).

Algoritme-modeller: Disse udgør de "abstrakte konceptuelle rammer, der bruges til at forstå og analysere algoritmernes struktur og adfærd" (Doe & Roe, 2020, frit oversat). Centralt heri er overvejelser om tidskompleksitet, rumkompleksitet og algoritmisk effektivitet.

Et eksempel: Del-og-hersk-modellen, et udbredt paradigme inden for algoritmeteori, opdeler et problem i mindre, håndterbare delproblemer og løser dem uafhængigt af hinanden, før resultaterne syntetiseres for at finde løsningen på det oprindelige problem (Tanenbaum & Austin, 2012).

Algoritmers begrænsninger: Dette refererer til de "begrænsninger eller iboende svagheder, der kan hæmme en algoritmes ydeevne, nøjagtighed eller

anvendelighed" (Li et al., 2017, frit oversat). Disse begrænsninger kan opstå på grund af forskellige faktorer, såsom problemkompleksitet eller datakvalitet.
Et eksempel: *Travelling Salesman*-problemet (TSP) er et eksempel på, at det er svært at bruge computerberegning i scenarier med store datasæt. Selvom der findes heuristiske metoder til approksimation, er det stadig en kompleks opgave at finde en optimal løsning (Cook, 2016).
Ved at dykke ned i de grundlæggende aspekter af algoritmer, deres modeller og begrænsninger bliver de studerende rustet til at konceptualisere og konstruere mere effektive løsninger til en række problemer, baseret på teori og praktisk anvendelse.

1.2 Forstå algoritmers rolle i forskellige sektorer og de potentielle faldgruber, der opstår ved brugen af dem

At få en nuanceret forståelse af algoritmernes rolle på tværs af sektorer og samtidig erkende potentielle faldgruber er afgørende for at fremme et holistisk syn på de implikationer og det ansvar, der er forbundet med implementering af algoritmer (Williamson, 2018).

Algoritmernes rolle i forskellige sektorer: Som fremhævet af Parker et al. (2016, frit oversat), "har algoritmer gennemsyret alle sektorer og revolutioneret processer og beslutningstagning". Eksempler på nogle vigtige sektorer, hvor algoritmisk indflydelse har særlig betydning, er:

a. Finansverdenen
b. Sundhedssektoren
c. E-handel
d. Transport

Potentielle faldgruber ved brug af algoritmer: På trods af deres mange fordele kan brug af algoritmer indebære betydelige faldgruber såsom bias og etiske dilemmaer, hvilket giver anledning til bekymring om "gennemsigtighed og potentiel overdreven afhængighed af automatisering" (O'Neil, 2016, frit oversat).

a. Bias og diskrimination
b. Bekymring om opretholdelse af privatlivets fred
c. Mangel på gennemsigtighed og ansvarlighed
d. Overdreven afhængighed af automatisering

At forstå algoritmernes mangefacetterede rolle og potentielle farer giver de studerende mulighed for at arbejde med dem på en ansvarlig og etisk korrekt

måde, hvilket fremmer kritisk tænkning og etisk forvaltning i vores digitale tidsalder.

1.3 Forstå algoritmisk kompleksitet, optimering og begrænsninger ved algoritmisk beslutningstagning

Hvis man får en dybere forståelse af algoritmisk kompleksitet, optimeringsstrategier og de begrænsninger, der ligger i algoritmiske beslutningsprocesser, kan det fremme en større forståelse af nuancerne ved algoritmer og deres udfordringer (Zhang, 2019).

Algoritmisk kompleksitet: Dette begreb er grundlæggende set en repræsentation af de beregningsressourcer, som en algoritme kræver, typisk udtrykt ved hjælp af *Big O-notation*, der fungerer som en "matematisk notation, der beskriver en funktions begrænsende opførsel" (Cormen et al., 2009, frit oversat).

Et eksempel: At skelne mellem lineær og logaritmisk tidskompleksitet giver et indblik i effektiviteten og ineffektiviteten af algoritmer som lineær søgning og binær søgning (Sedgewick & Wayne, 2011).

Optimering: Som påpeget af Karp (2010, frit oversat) er "optimering kunsten at gøre en algoritme mere effektiv", hvilket omfatter strategier til at minimere kompleksiteten og forbedre performance gennem forskellige innovative tilgange.

Et eksempel: Implementering af mere effektive sorteringsalgoritmer, såsom *quicksort*, illustrerer det transformerende potentiale af optimering i beregningsprocesser (Bentley, 1999).

Begrænsninger ved algoritmisk beslutningstagning: På trods af fordelene er der også væsentlige begrænsninger, der ofte er knyttet til "datakvalitet og etiske overvejelser, hvilket nogle gange fører til utilsigtede konsekvenser" (Diaz & Sonsini, 2016, frit oversat).

Et eksempel: Inden for strafferetsplejen har algoritmiske forudsigelser skabt debat om retfærdighed og potentielle bias, hvilket understreger det komplekse etiske landskab ved algoritmisk implementering (Angwin et al., 2016).

Gennem en omfattende udforskning af algoritmisk kompleksitet, optimeringsstrategier og de iboende begrænsninger i algoritmisk beslutningstagning bliver de studerende i stand til at navigere i det komplekse område inden for algoritmedesign og implementering fra et informeret og kritisk perspektiv.

CU2 – Data-fairness og bias i AI EQF6 LÆRINGSUDBYTTE		
	Efter endt læringsforløb vil deltageren være i stand til at	
VIDEN	**FÆRDIGHEDER**	**HOLDNINGER**
Teoretisk/faglig viden i: CUK2.1. Forstå begreberne retfærdighed og bias i AI-systemer. CUK2.2. Kende til de forskellige kilder til og manifestationer af bias i algoritmiske processer. CUK2.3. Lære metoder til at identificere og måle bias i data. CUK2.4. Forstå strategier til at håndtere og afbøde effekterne af bias for at sikre mere retfærdige resultater.	Færdighedsmål 1: Forstå begreberne retfærdighed og bias i AI-systemer: Færdighed 1.1: Anvende begreberne retfærdighed og bias til at vurdere de etiske implikationer af AI-drevne løsninger i forskellige sammenhænge. Færdighed 1.2: Integrere overvejelser om retfærdighed og bias i design og udvikling af AI-systemer, hvor der lægges vægt på etisk og social ansvarlighed. Færdighedsmål 2: Kende til de forskellige kilder til og manifestationer af bias i algoritmiske processer: Færdighed 2.1: Bruge analytiske teknikker til at opdage og diagnosticere forskellige typer af bias i algoritmiske processer, såsom datadrevet bias, modeldrevet bias og menneskedrevet bias. Færdighed 2.2: Vurdere de potentielle kilder til bias i forskellige stadier af AI-udviklingsprocessen, fra dataindsamling til modelimplementering. Færdighedsmål 3: Lære metoder til at identificere og måle bias i data: Færdighed 3.1: Anvende forskellige metoder og værktøjer til at identificere, måle og kvantificere bias i datasæt og algoritmiske outputs. Færdighed 3.2: Løbende overvåge og vurdere AI-systemer med henblik på at opdage potentielle bias og utilsigtede konsekvenser samt justere afhjælpningsstrategier efter behov. Færdighedsmål 4: Forstå strategier til at håndtere og afbøde effekterne af bias for at sikre mere retfærdige resultater: Færdighed 4.1: Udvikle og implementere strategier til at adressere og afbøde bias i AI-systemer under hensyntagen til	Holdningsmål 1: Forstå begreberne retfærdighed og bias i AI-systemer: Holdning 1.1: Anerkende vigtigheden af retfærdighed og lighed i AI-systemer og fremme etiske overvejelser i systemernes udvikling og anvendelse. Holdningsmål 2: Kende til de forskellige kilder til og manifestationer af bias i algoritmiske processer. Holdningsmål 2.1: Have en proaktiv tilgang til at identificere og adressere potentielle bias i hele AI-udviklingsprocessen. Holdningsmål 3: Lære metoder til at identificere og måle bias i data: Holdningsmål 3.1: Forstå betydningen af grundig og løbende identifikation og måling af bias for at sikre ansvarlig udvikling og brug af AI-systemer. Holdningsmål 4: Forstå strategier til at adressere og afbøde effekterne af bias for at sikre mere retfærdige resultater: Holdningsmål 4.1: Forpligte sig til at adressere og afbøde bias i AI-systemer for at fremme retfærdige resultater for alle brugere. Holdningsmål 4.2: Fremme en kultur med samarbejde og løbende læring, hvor man involverer sig med forskellige interessenter for at sikre en ansvarlig og etisk brug af AI-systemer. Holdningsmål 4.3: Forstå vigtigheden af at holde sig informeret om den nyeste forskning, de nyeste tendenser og best practice inden for AI-etik, retfærdighed og bias-reducering for løbende at forbedre den faglige kompetence.

	kompromiserne mellem nøjagtighed, retfærdighed og andre performance-målinger. Færdighed 4.2: Evaluere effektiviteten af teknikker til afhjælpning af bias for at opnå mere retfærdige resultater i AI-applikationer. Færdighed 4.3: Samarbejde med forskellige interessenter, herunder dataforskere, ingeniører og andre eksperter på området, for at sikre en ansvarlig og etisk brug af AI-systemer. Færdighed 4.4: Kommunikere udfordringer og løsninger relateret til retfærdighed og bias i AI-systemer til både tekniske og ikke-tekniske målgrupper. Færdighed 4.5: Holde sig ajour om den nyeste forskning, tendenser og best practice inden for AI-etik, retfærdighed og bias-begrænsning for løbende at forbedre den faglige kompetence.	
KOMPETENCE/ER I DIGCOMP2.2., ENTRECOMP OG GREENCOMP-FRAMEWORKS, SOM CU2 ER FORBUNDET MED - Er bevidst om potentielle informationsbias forårsaget af forskellige faktorer (f.eks. data, algoritmer, redaktionelle valg, censur, ens egne personlige begrænsninger). (DIGCOMP2.2.) - Er bevidst om, at AI-algoritmer måske ikke er konfigureret til kun at give den information, som brugeren ønsker; de kan også indeholde et kommercielt eller politisk budskab (f.eks. for at opmuntre brugerne til at blive på siden, se eller købe noget bestemt, tilkendegive en bestemt holdning). Dette kan også have negative konsekvenser (f.eks. reproduktion af stereotyper, deling af misinformation). (DIGCOMP2.2.) - Forstår, at mens anvendelsen af AI-systemer på mange områder normalt er ukontroversiel (f.eks. AI, der hjælper med at bekæmpe	**FÆRDIGHEDER I DIGCOMP2.2., ENTRECOMP OG GREENCOMP RAMMERNE, SOM CU2 ER FORBUNDET MED** - Kunne forstå, at nogle AI-algoritmer kan forstærke eksisterende synspunkter i digitale miljøer ved at skabe "ekkokamre" eller "filterbobler" (f.eks. hvis en række sociale medier favoriserer en bestemt politisk ideologi, kan yderligere anbefalinger af ideologien forstærke den uden at give den modspil). (DIGCOMP2.2.) - Ved, hvordan man ændrer brugerkonfigurationer (f.eks. i apps, software, digitale platforme) for at aktivere, forhindre eller moderere AI-systemets sporing, indsamling eller analyse af data (f.eks. ved ikke at tillade mobiltelefonen at spore brugerens placering). (DIGCOMP2.2.)	**HOLDNINGER I DIGCOMP2.2., ENTRECOMP OG GREENCOMP-RAMMEVÆRKERNE, SOM CU2 ER FORBUNDET MED** - Kan identificere både de positive og negative konsekvenser af brugen af alle data (indsamling, kodning og behandling), men især personlige data, af AI-drevne digitale teknologier som apps og onlinetjenester. (DIGCOMP2.2.)

klimaforandringer), kan AI, der direkte interagerer med mennesker og træffer beslutninger om deres liv, ofte være kontroversiel (f.eks. CV-sorteringssoftware til rekrutteringsprocedurer, bedømmelse af eksamener, der kan afgøre optagelse på en uddannelse) (DIGCOMP2.2). - Er klar over, at AI-systemer indsamler og behandler flere typer brugerdata (f.eks. personlige data, adfærdsdata og kontekstuelle data) for at skabe brugerprofiler, som derefter bruges til f.eks. at forudsige, hvad brugeren måske vil se eller gøre næste gang (f.eks. vise bestemte reklamer, give bestemte anbefalinger, tilbyde forskellige tjenester). (DIGCOMP2.2.) - Ved, at etiske begreber og retfærdighed for nuværende og fremtidige generationer er relateret til beskyttelse af naturen (GREENCOMP)		

CU2 BESKRIVELSE AF VIDENSUDBYTTE

Hovedformålet med denne CU er at give studerende på EQF6-niveau en dyb forståelse af kompleksiteten i dataretfærdighed og bias inden for kunstig intelligens, hvilket fremmer en etisk og inkluderende praksis inden for dette område.

2.1. Forstå begreberne retfærdighed og bias i AI-systemer

De studerende vil blive uddannet til at få en nuanceret forståelse af det teoretiske grundlag for retfærdighed og bias inden for AI-systemer. Baseret på grundlæggende teorier vil de studerende studere de moralske og etiske implikationer nærmere, der omgiver AI-teknologier. Som Mittelstadt et al. (2016, frit oversat) bemærker, er denne viden afgørende for at "identificere og forstå de etiske implikationer af algoritmisk beslutningstagning".

Forklaring:

Teoretiske rammer: De studerende vil dykke ned i teorier og rammer, hvor der bliver set nærmere på begreberne retfærdighed og bias i AI.

Etiske implikationer: En analyse af de etiske implikationer af disse begreber i virkelige scenarier vil blive gennemført, hvilket fremmer en kritisk tilgang til AI-systemudvikling.

Eksempel: Detaljerede casestudier, der viser, hvordan algoritmiske bias manifesterer sig i forskellige sektorer såsom sundhedsvæsenet, finansverdenen og retssystemet.

2.2. Kende til de forskellige kilder til og manifestationer af bias i algoritmiske processer

I dette modul vil de studerende blive trænet i at identificere de potentielle kilder til og manifestationer af bias via en undersøgende tilgang, som Barocas et al. (2019) også er inde på, hvor Barocas et al. opfordrer kraftigt til, at man kan identificere bias fra "datagenerering til inferens".

Forklaring:

Bias ved datagenerering: Deltagerne vil kritisk undersøge, hvordan bias kan opstå i datagenereringsprocessen og lære strategier til at forhindre det.

Bias i inferens: Deltagerne vil undersøge, hvordan bias kan vise sig under algoritmiske inferenser og de potentielle konsekvenser af sådanne bias.

Eksempel: Analyse af scenarier fra den virkelige verden, hvor bias er blevet identificeret i forskellige faser af algoritmiske processer, og en kritisk evaluering af de foranstaltninger, der er truffet for at håndtere dem.

2.3. Lære metoder til at identificere og måle bias i data

Denne del lægger vægt på at tilegne sig praktiske færdigheder i at identificere og måle bias i datasæt. Ifølge Danks og London (2017, frit oversat) er brugen af robuste metoder afgørende for at opdage og afbøde "algoritmiske bias, der kan have uberettigede differentierede virkninger".

Forklaring:

Analytiske teknikker: De studerende vil blive fortrolige med forskellige analytiske teknikker og værktøjer til effektivt at identificere og måle bias i data.

Kritisk evaluering: De studerende vil blive opfordret til kritisk at evaluere eksisterende algoritmer og foreslå forbedringer for at mindske bias.

Eksempel: Workshops, hvor de studerende får praktisk erfaring med at anvende forskellige analytiske teknikker til at identificere og måle bias i datasæt.

2.4. Forstå strategier til at håndtere og afbøde effekterne af bias for at sikre mere retfærdige resultater

For at fremme en proaktiv tilgang lægger denne del vægt på strategier til at adressere og afbøde bias i AI-systemer. Som formuleret af Benjamin (2019) er disse strategier afgørende for at sikre skabelsen af AI-teknologier, der fremmer "et mere retfærdigt og ligeværdigt samfund".

Forklaring:

Afhjælpende strategier: Deltagerne vil undersøge forskellige strategier og teknikker, der kan bruges til at håndtere og reducere bias i AI-systemer.
Etiske tilgange: Denne del lægger vægt på udviklingen af etiske tilgange til AI-udvikling, der fremmer inklusivitet og retfærdighed.
Et eksempel: Udvikling af et projekt, hvor de studerende formulerer og implementerer strategier til at håndtere bias i AI-systemer, hvilket fremmer mere retfærdige resultater.

Ved omhyggeligt at navigere gennem de indviklede forhold omkring dataretfærdighed og bias i kunstig intelligens vil de studerende blive klædt på til at fremme ansvarlig og inkluderende AI-udvikling og få rollen som banebrydende eksperter inden for kunstig intelligens.

CU3 – Privatlivets fred og bekvemmelighed i AI EQF6 **LÆRINGSUDBYTTE**		
	Efter endt læringsforløb vil deltageren være i stand til at	
VIDEN	**FÆRDIGHEDER**	**HOLDNINGER**
Teoretisk/faglig viden i: CUK3.1. Forstå kompromiserne mellem privatlivets fred og teknologiens bekvemmelighed i AI-applikationer. CUK3.2. Forstå udfordringerne ved at skulle opretholde privatlivets fred hos brugerne og samtidig levere personaliserede og effektive services. CUK3.3. Kende til eksisterende og nye teknologier, der er designet til at skabe balance mellem privatlivets fred og teknologiens bekvemmelighed i AI-systemer.	Færdighedsmål 1: Forstå kompromiserne mellem privatlivets fred og teknologiens bekvemmelighed i AI-applikationer. 1.1. Analysere casestudier om AI-applikationer for at finde frem til en balance mellem privatlivets fred og teknologiens bekvemmelighed. 1.2. Kritisk evaluere de etiske overvejelser ved design af AI-systemer med hensyn til privatlivets fred og teknologiens bekvemmelighed. 1.3. Udvikle strategier til at minimere potentielle privatlivsrisici i AI-applikationer, samtidig med at brugervenligheden opretholdes. Færdighedsmål 2: Forstå udfordringerne ved at skulle opretholde privatlivets fred hos brugerne og samtidig levere personaliserede og effektive services. 2.1. Identificere de vigtigste udfordringer, som AI-udviklere står over for, når de skal sikre brugernes privatliv og samtidig levere personaliserede services. 2.2. Vurdere potentielle risici og sårbarheder i AI-systemer, der kan kompromittere brugernes privatliv. 2.3. Foreslå løsninger til at afbøde udfordringerne vedrørende privatlivets fred	Holdningsmål 1: Forstå kompromiserne mellem privatlivets fred og teknologiens bekvemmelighed i AI-applikationer. 1.1. Udvikle ansvarsfølelse med hensyn til at respektere privatlivets fred i AI-applikationer. 1.2. Udvikle en etisk tankegang med hensyn til at skabe balance mellem privatlivets fred og teknologiens bekvemmelighed i AI-systemdesign. 1.3. Forstå vigtigheden af brugernes tillid under udformningen af AI-applikationer, der respekterer privatlivets fred. Holdningsmål 2: Forstå udfordringerne ved at skulle opretholde privatlivets fred hos brugerne og samtidig levere personaliserede og effektive services. 2.1. Udvise empati over for brugernes bekymringer om privatlivets fred i forbindelse med personaliserede AI-services. 2.2. Fremme et engagement i at håndtere privatlivsudfordringer i AI-udvikling.

	og samtidig bibeholde de personaliserede services' effektivitet i AI-applikationer. Færdighedsmål 3: Kende til eksisterende og nye teknologier, der er designet til at skabe balance mellem privatlivets fred og teknologiens bekvemmelighed i AI-systemer. 3.1. Undersøge og analysere eksisterende teknologier og deres tilgange til at løse problemer med privatlivets fred og teknologiens bekvemmelighed i AI-applikationer. 3.2. Evaluere nye teknologiers styrker og svagheder med hensyn til at skabe balance mellem privatlivets fred og teknologiens bekvemmelighed i AI-systemer. 3.3. Designe og foreslå nye løsninger eller forbedringer af eksisterende teknologier for at forbedre balancen mellem privatlivets fred og teknologiens bekvemmelighed i AI-applikationer.	2.3. Forstå vigtigheden af kontinuerlig forbedring af privatlivsbevarende teknikker til personaliserede AI-services. Holdningsmål 3: Kende til eksisterende og nye teknologier, der er designet til at skabe balance mellem privatlivets fred og teknologiens bekvemmelighed i AI-systemer. 3.1. Være nysgerrig og åben over for at lære om nye teknologier og metoder inden for privatlivsbevarende AI-systemer. 3.2. Forstå AI-feltets innovative natur og dets potentiale til at forbedre privatlivets fred og teknologiens bekvemmelighed. 3.3. Have en samarbejdsorienteret tilgang til problemløsning i erkendelse af, at det er et fælles ansvar blandt interessenter at løse problemer med privatlivets fred og teknologiens bekvemmelighed i AI.
KOMPETENCE/ER I DIGCOMP2.2., ENTRECOMP OG GREENCOMP-FRAMEWORKS, SOM CU3 ER FORBUNDET MED - Bevidst om, at sensorer, der bruges i mange digitale teknologier og applikationer (f.eks. ansigtssporingskameraer, virtuelle assistenter, kropsbåren teknologi, mobiltelefoner, smarte enheder), genererer store mængder data, herunder personlige data, der kan bruges til at træne et AI-system. (DIGCOMP2.2.) - Ved, at indsamlet og behandlet data, f.eks. af onlinesystemer, kan bruges til at genkende mønstre (f.eks. repetitioner) i nye data (f.eks. andre billeder, lyde, museklik, onlineadfærd) for yderligere at optimere og personliggøre online-tjenester (f.eks. reklamer). (DIGCOMP2.2.) - Bevidst om, at alt, hvad man deler offentligt online (f.eks. billeder, videoer, lyde), kan bruges til at træne AI-systemer. For eksempel kan kommercielle softwarevirksomheder, der	**FÆRDIGHEDER I DIGCOMP2.2., ENTRECOMP OG GREENCOMP FRAMEWORKS, SOM CU3 ER FORBUNDET MED** - Kan bruge dataværktøjer (f.eks. databaser, datamining, analysesoftware), som er designet til at håndtere og organisere kompleks information, til at understøtte beslutningstagning og problemløsning (DIGCOMP2.2.) - Ved, hvordan man opdager, om man kommunikerer med et menneske eller en AI-baseret samtaleagent (f.eks. når man bruger tekst- eller stemmebaserede chatbots). (DIGCOMP2.2.) - Kan inkorporere AI-redigeret/manipuleret digitalt indhold i sit eget arbejde (f.eks. inkorporere AI-genererede melodier i ens egen musikalske komposition). Denne brug af AI kan være kontroversiel, da den rejser spørgsmål om AI's rolle i kunstværker, for eksempel hvem der skal krediteres for det. (DIGCOMP2.2.) - Udtryk for bæredygtighedsværdier: 1.2 Understøtte retfærdighed - at understøtte lighed og retfærdighed for	**HOLDNINGER I DIGCOMP2.2., ENTRECOMP OG GREENCOMP FRAMEWORKS, SOM CU3 ER FORBUNDET MED** - Overvejer gennemsigtighed, når data manipuleres og præsenteres, for at sikre deres pålidelighed, og opdager, hvis data vises med skjulte motiver (f.eks. uetisk, profit, manipulation) eller på vildledende måder (DIGCOMP2.2.) - Er åben for AI-systemer, der støtter mennesker i at træffe informerede beslutninger i overensstemmelse med deres mål (f.eks. brugere, der aktivt beslutter, om de vil handle på baggrund af en anbefaling eller ej) (DIGCOMP2.2.) - Overvejer etiske aspekter (herunder, men ikke begrænset til, menneskelig udførelse og kontrol, gennemsigtighed, ikke-diskrimination, tilgængelighed samt bias og retfærdighed) som en af de vigtigste søjler, når man udvikler eller implementerer AI-systemer. (DIGCOMP2.2.)

udvikler AI-ansigtsgenkendelsessystemer, bruge personlige billeder, der deles online (f.eks. familiefotografier), til at træne og forbedre softwarens evne til automatisk at genkende personer på andre billeder, hvilket måske ikke er ønskeligt (f.eks. kan det være et brud på privatlivets fred) (DIGCOMP2.2.) - Ved, at AI-systemer kan bruges til automatisk at skabe digitalt indhold (f.eks. tekster, nyheder, essays, tweets, musik, billeder) ved at bruge eksisterende digitalt indhold som kilde. Sådant indhold kan være svært at skelne fra menneskelige kreationer. (DIGCOMP2.2.) - Ved, at behandling af personoplysninger er underlagt lokale bestemmelser som EU's Databeskyttelsesforordning (GDPR) (f.eks. er stemmeinteraktioner med en virtuel assistent personoplysninger i henhold til GDPR og kan udsætte brugerne for visse databeskyttelses-, privatlivs- og sikkerhedsrisici). (DIGCOMP2.2.) - Er opmærksom på, at visse aktiviteter (f.eks. træning af AI og produktion af kryptovalutaer som Bitcoin) er ressourcekrævende processer med hensyn til data og computerkraft. Derfor kan energiforbruget være højt, hvilket også kan have en stor miljøpåvirkning. (DIGCOMP2.2.) - Bevidst om, at AI er et produkt af menneskelig intelligens og beslutningstagning (dvs. mennesker vælger, renser og koder dataene, de designer algoritmerne, træner modellerne og organiserer og anvender menneskelige værdier på outputtet) og eksisterer derfor ikke uafhængigt af mennesker (DIGCOMP2.2.)	nuværende og fremtidige generationer og lære af tidligere generationer med henblik på bæredygtighed. 1.3 Tage vare på naturen - at anerkende, at mennesker er en del af naturen, og at respektere andre arters og naturens egne behov for og rettigheder til at genoprette og regenerere sunde og modstandsdygtige økosystemer. (GreenComp)	- Afvejer fordele og risici, før man tillader tredjeparter at behandle personlige data (forstår f.eks., at en stemmeassistent på en smartphone, der bruges til at give kommandoer til en robotstøvsuger, kan give tredjeparter - virksomheder, regeringer, cyberkriminelle - adgang til dataene). (DIGCOMP2.2.) - Overvejer de etiske konsekvenser af AI-systemer i hele deres livscyklus: de omfatter både miljøpåvirkningen (dvs. de miljømæssige konsekvenser af produktionen af digitale enheder og tjenester) og samfundsmæssige konsekvenser, f.eks. platformisering af arbejde og algoritmisk styring, der kan undertrykke arbejdstagernes privatliv eller rettigheder; brugen af billig arbejdskraft til rubricering af billeder for at træne AI-systemer. (DIGCOMP2.2.)

CU3 BESKRIVELSE AF VIDENSRESULTATER

Hovedformålet med denne CU er at give studerende på EQF6-niveau en dyb forståelse af det komplekse samspil mellem opretholdelse af privatlivets fred og teknologiens bekvemmelighed inden for AI, udvikle en kritisk bevidsthed og de færdigheder, der er nødvendige for at navigere og være innovativ på en ansvarlig måde inden for dette felt.

3.1. Forstå dikotomien om privatlivets fred og teknologiens bekvemmelighed i AI-applikationer

De studerende vil fordybe sig i et intensivt studie af den fine balance mellem privatlivets fred og teknologiens bekvemmelighed, som er to centrale aspekter, der er udslagsgivende for, i hvilken grad AI-systemer accepteres og integreres i samfundet. Ved at følge Zuboffs (2019) diskurs, som belyser den "overvågningskapitalisme", der ofte skygger for teknologiske fremskridt, vil de studerende analysere casestudier og forskellige perspektiver, der styrer denne dikotomi.

Forklaring:

Historisk perspektiv: Et kig på, hvordan vægten mellem privatlivets fred og teknologiens bekvemmelighed er tippet gennem tiden, hvor AI-teknologier er blevet udviklet.

Filosofisk underbygning: Deltagerne vil studere de filosofiske debatter om privatlivets fred og teknologiens bekvemmelighed i en digital tidsalder.

Et eksempel: Analyse af scenarier, hvor teknologiske fremskridt har overtrådt normerne for privatlivets fred i jagten på bekvemmelighed og effektivitet.

3.2. Identificere og analysere potentielle brud på privatlivets fred i AI-implementeringer

I dette modul vil de studerende opnå færdighederne til at identificere og kritisk analysere potentielle brud på privatlivets fred i forskellige AI-implementeringer. Dette er i tråd med Pasquales (2015) forskning, som insisterer på nødvendigheden af et "black box-samfund", hvor uigennemsigtigheden af algoritmiske processer bør undersøges for at sikre privatlivets fred.

Forklaring:

Analytiske færdigheder: Udvikling af analytiske færdigheder til at identificere potentielle brud på privatlivets fred i AI-applikationer.

Juridiske perspektiver: Et dybdegående studie af de juridiske rammer for privatlivets fred i forbindelse med AI.

Eksempler: Casestudier, hvor bemærkelsesværdige brud på privatlivets fred analyseres, og hvor de erfaringer, der er gjort i den forbindelse, undersøges samt de foranstaltninger, der er indført som reaktion på bruddet.

3.3. Udforske teknologiske fremskridt, der sigter mod at balancere privatlivets fred og teknologiens bekvemmelighed

Denne del har til formål at gøre deltagerne bekendt med de seneste teknologiske fremskridt, hvorigennem man forsøger at opretholde en harmonisk balance mellem privatlivets fred og teknologiens bekvemmelighed. Deltagerne vil blive præsenteret for Nissenbaum (2009), som er fortaler for "privatlivets fred i kontekst" som en vej til at forstå den nuancerede karakter af eventuelle problemer med privatlivets fred i AI-landskabet.

Forklaring:

Nye teknologier: En detaljeret undersøgelse af nye teknologier, der søger at skabe balance mellem privatlivets fred og teknologiens bekvemmelighed.

Innovative tilgange: Analyse af innovative tilgange, som forskellige sektorer har anvendt for at sikre privatlivets fred uden at gå på kompromis med bekvemmeligheden.

Et eksempel: Undersøgelse af nye teknologiske løsninger som f.eks. *differential privacy*, der gør det muligt at udnytte og benytte data og samtidig bevare privatlivets fred.

3.4. Udvikle strategier til at fremme privatlivets fred uden at gå på kompromis med teknologiens bekvemmelighed i AI-systemer

For at give deltagerne evnen til at kunne bidrage aktivt til emnet, fokuserer denne del på at udvikle strategier, der fremmer privatlivets fred uden at gå på kompromis med bekvemmeligheden i AI-systemer. Deltagerne vil beskæftige sig med værker af forskere som Cavoukian (2010), der foreslog konceptet "Privacy by Design" som en fremsynet tilgang til at integrere privatlivets fred i udviklingen af AI-systemer.

Forklaring:

Strategisk udvikling: Anspore de studerende til at udvikle strategier, der integrerer beskyttelse af privatlivets fred uden at gå på kompromis med den effektivitet og bekvemmelighed, som AI tilbyder.

Etiske overvejelser: Fokus på de etiske overvejelser, der styrer udviklingen af AI-teknologier, specielt med hensyn til opretholdelse af privatlivets fred.

Eksempel: De studerende vil deltage i workshops for at udvikle AI-løsninger, der også har beskyttelse af privatlivets fred som en kernefunktion, hvilket fremmer ansvarlig innovation.

Ved at rejse igennem det mangefacetterede landskab, hvor privatlivets fred og teknologiens bekvemmelighed i AI skal leve side om side, vil de studerende blive rustet til kritisk at analysere, innovere og lede udviklingen af AI-teknologier, der respekterer principperne om privatlivets fred og samtidig tilbyder overlegen teknologisk bekvemmelighed.

CU4 – AI-etik, en praktisk tilgang EQF6 **LÆRINGSUDBYTTE**		
	Efter endt læringsforløb vil deltageren være i stand til at	
VIDEN	**FÆRDIGHEDER**	**HOLDNINGER**
Teoretisk/faglig viden i: CUK4.1. Være bevidst om de etiske principper i udvikling og implementering af AI. CUK4.2. Kende til forskellige etiske rammer og retningslinjer samt deres anvendelse i virkelige scenarier, hvor AI-systemer indgår. CUK4.3. Forstå vigtigheden af gennemsigtighed, ansvarlighed og inddragelse af interessenter i den etiske AI-udviklingsproces.	Færdighedsmål 1: Anvende etiske principper i AI-udvikling og -implementering gennem praktiske tilgange. 1.1. Anvende etiske principper til design, udvikling og implementering af AI-systemer for at minimere algoritmisk bias. 1.2. Evaluere AI-systemer i forhold til etiske retningslinjer og principper for at identificere potentielle bias og etiske problemstillinger. 1.3. Udvikle strategier til at løse etiske problemer i AI-applikationer, herunder eventuelle konflikter mellem principper og kompromiser. Færdighedsmål 2: Analysere og evaluere forskellige etiske rammer og retningslinjer i virkelige situationer, hvor AI-systemer indgår. 2.1. Analysere forskellige etiske rammer og retningslinjer, der er relevante for AI og algoritmisk bias.	Holdningsmål 1: En etisk tankegang i AI-udvikling og -implementering. 1.1. Udvikle en følelse af ansvar for at anvende etiske principper i AI-udvikling med henblik på at minimere algoritmisk bias. 1.2. Fremme en etisk tankegang, hvori man værdsætter fairness, inklusivitet og retfærdighed i AI-applikationer. 1.3. Forstå vigtigheden af at adressere etiske bekymringer proaktivt for at forhindre utilsigtede konsekvenser af AI-systemer. Holdningsmål 2: Nysgerrighed og åbenhed over for at lære om forskellige etiske rammer og retningslinjer i virkelige situationer, hvor AI-systemer indgår. 2.1. Vise nysgerrighed og åbenhed over for at lære om forskellige etiske rammer og deres relevans for AI-udvikling. 2.2. Lægge vægt på vigtigheden af etiske retningslinjer i udformningen og implementeringen af AI-systemer.

	2.2. Anvende etiske rammer på virkelige AI-scenarier for at vurdere de etiske dimensioner af AI-systemer. 2.3. Sammenligne og kontrastere forskellige etiske rammer for at vurdere deres egnethed til at håndtere algoritmisk bias i specifikke AI-applikationer. Færdighedsmål 3: Designe og udvikle etiske AI-systemer ved at fremme gennemsigtighed, ansvarlighed og engagement fra interessenter. 3.1. Designe AI-systemer, der fremmer gennemsigtighed, så brugere og interessenter kan forstå beslutningsprocessen. 3.2. Implementere ansvarligheds-foranstaltninger for at sikre, at AI-udviklere og organisationer tager ansvar for konsekvenserne af deres AI-systemer. 3.3. Fremme interessenternes engagement i AI-udviklingen, herunder at indhente input med forskellige perspektiver og tage hensyn til behovene hos forskellige samfundsgrupper, der er berørt af AI-systemer.	2.3. Anvende en reflekterende tilgang til at evaluere og forfine AI-applikationer baseret på etiske overvejelser. Holdningsmål 3: Opmuntrende og entusiastisk tankegang, der fremmer gennemsigtighed, ansvarlighed og interessenters engagement i etisk AI-udvikling. 3.1. Udvikle en forpligtelse til at støtte gennemsigtighed i AI-systemer, så brugere og interessenter kan forstå og have tillid til AI-beslutningstagning. 3.2. Lægge vægt på vigtigheden af ansvarlighed i AI-udvikling og forstå behovet for, at organisationer tager ansvar for konsekvenserne af deres AI-systemer. 3.3. Tilskynde til aktiv inddragelse af interessenter og samarbejde i udviklingen af AI, og forstå værdien af forskellige perspektiver og inddragelse af berørte samfundsgrupper.
KOMPETENCE/ER I DIGCOMP2.2., ENTRECOMP OG GREENCOMP-FRAMEWORKS, SOM CU4 ER FORBUNDET MED - Ved, at AI i sig selv hverken er godt eller dårligt. Det, der afgør, om resultaterne af et AI-system er positive eller negative for samfundet, er, hvordan AI-systemet er designet og brugt, af hvem og til hvilke formål (AI) (DigiComp 2.2.) - Ved, at behandlingen af personoplysninger er underlagt regionale regler såsom EU's Databeskyttelsesforordning (GDPR) (f.eks. er taleinteraktioner med en virtuel assistent persondata i forhold til GDPR og kan udsætte brugere for forskellige former for risici med hensyn til databeskyttelse, privatliv og sikkerhed). (AI) (DigiComp 2.2.)	**FÆRDIGHEDER I DIGCOMP2.2.-, ENTRECOMP- OG GREENCOMP-RAMMERNE, SOM CU4 ER FORBUNDET MED** - Ved, hvordan man identificerer områder, hvor AI kan være en fordel i forskellige hverdagssituationer. For eksempel kan AI i sundhedsvæsenet bidrage til tidlig diagnosticering, mens det i landbruget kan bruges til at opdage skadedyrsangreb. (AI). (DigiComp 2.2.) - At samarbejde: Holde sammen, arbejde sammen og netværke. Samarbejde med andre for at udvikle ideer og omsætte dem til handling. Netværke. Løse konflikter og se konkurrence i øjnene på en positiv måde, når det er nødvendigt. (EntreComp) - Udtryk for bæredygtighedsværdier: 1.2 Understøtte retfærdighed - at understøtte lighed og retfærdighed for nuværende og fremtidige generationer og	**HOLDNINGER I DIGCOMP2.2., ENTRECOMP OG GREENCOMP FRAMEWORKS, SOM CU4 ER FORBUNDET MED** - Er parat til at overveje etiske spørgsmål relateret til AI-systemer (f.eks. i hvilke sammenhænge, såsom strafudmåling af kriminelle, bør AI-anbefalinger ikke bruges uden menneskelig indgriben?) (AI) (DigiComp 2.2.) - Er åben for AI-systemer, der støtter mennesker i at træffe informerede beslutninger i overensstemmelse med deres mål (f.eks. brugere, der aktivt beslutter, om de vil handle på baggrund af en anbefaling eller ej) (DIGCOMP2.2.) - Betragter etik (herunder, men ikke begrænset til, menneskelig udførelse og kontrol, gennemsigtighed, ikke-diskriminering, tilgængelighed og bias samt retfærdighed) som en af grundpillerne ved udvikling eller implementering af AI-systemer. (AI) (DigiComp 2.2.) - Er åben for at engagere sig i samarbejdsprocesser for at co-designe og co-skabe nye produkter og services baseret på AI-systemer for at understøtte og øge

	lære af tidligere generationer med henblik på bæredygtighed. 1.3 Tage vare på naturen - at anerkende, at mennesker er en del af naturen, og at respektere andre arters og naturens egne behov for og rettigheder til at genoprette og regenerere sunde og modstandsdygtige økosystemer. (GreenComp)	borgernes deltagelse i samfundet. (AI) (DigiComp 2.2.) Kreativitet (f.eks. være nysgerrig og åben): Udvikle kreative og målrettede ideer. Udvikle flere ideer og muligheder for at skabe værdi, herunder bedre løsninger på eksisterende og nye udfordringer. Udforske og eksperimentere med innovative tilgange. Kombinere viden og ressourcer for at opnå resultater, som har værdi. (EntreComp) Mobilisering af andre: Inspirere og begejstre relevante interessenter. Få den nødvendige støtte til at opnå resultater, som har værdi. Udvise at man er i stand til effektiv kommunikation, overtalelse, forhandling og lederskab. (EntreComp)

CU4 BESKRIVELSE AF VIDENSRESULTATER

4.1. Være bevidst om de etiske principper i udvikling og implementering af AI

Være bevidst om de etiske principper i udvikling og implementering af AI. Anvende etiske principper til design, udvikling og implementering af AI-systemer for at minimere algoritmisk bias. Evaluere AI-systemer i forhold til etiske retningslinjer og principper for at identificere potentielle bias og etiske problemstillinger. Udvikle strategier til at løse etiske problemer i AI-applikationer, herunder potentielle konflikter mellem principper og kompromiser.

Etiske principper for AI: Etiske principper har til formål at skabe et grundlag for at få AI-systemer til at fungere til gavn for menneskeheden, enkeltpersoner, samfundet, miljøet og økosystemet samt for at forhindre skadelige virkninger. Principperne er styrende for design, udvikling, implementering og brug af AI-systemer på en måde, der er troværdig og sætter menneskelig værdighed, lighed mellem alle mennesker, bevarelse af miljø, biodiversitet og økosystemer, respekt for kulturel mangfoldighed og dataansvar i centrum (UNESCO, 2022). De etiske principper beskriver, hvad der forventes med hensyn til rigtigt og forkert samt andre etiske standarder. De etiske principper for AI henviser til normative begrænsninger for *"do's"* og *"don'ts"* ved brug af algoritmer i samfundet (Zhou et al., 2020). Ifølge *High-Level Expert Group on AI (AI HLEG)* omfatter sådanne principper for eksempel: respekt for menneskers autonomi, skadeforebyggelse, retfærdighed og forklarlighed (Europa-Kommissionen, 2019).

Praktisk anvendelse af etiske principper i AI. Når de etiske principper er identificeret, bør de omsættes til brugbare værktøjer og retningslinjer, så de kan præge AI-baseret innovation og støtte den praktiske anvendelse af etiske principper for AI. Værktøjer og retningslinjer for, hvordan man anvender etiske principper i design, implementering og udrulning af AI, er yderst vigtige. Etiske AI-algoritmer, -arkitekturer og -grænseflader følger etiske principper for AI. (Zhou et al., 2020.) Mange private, offentlige og almennyttige organisationer har i løbet af deres arbejde med at udvikle de etiske principper for AI identificeret nyttige tiltag, der kan hjælpe med at formulere og implementere deres foreslåede principper (se f.eks. UNESCO, 2021).

For eksempel skal de principper, der er skitseret af AI HLEG, omsættes til konkrete krav for at opnå troværdige AI-systemer. Disse krav gælder for forskellige interessenter, der deltager i AI-systemers livscyklus: udviklere, de der implementerer AI, samt slutbrugere, såvel som samfundet generelt. Forskellige grupper af interessenter spiller forskellige roller for at sikre, at kravene opfyldes: udviklere bør implementere og anvende kravene i design- og udviklingsprocesser; de, der implementerer AI, bør sikre, at de systemer, de bruger, og de produkter og services, de tilbyder, opfylder kravene; slutbrugere og samfundet generelt bør informeres om disse krav og være i stand til at anmode om, at de opretholdes. Nedenstående liste over krav er ikke udtømmende; den omfatter systemiske, individuelle og samfundsmæssige aspekter:

1. Menneskelig udførelse og kontrol (dvs. grundlæggende rettigheder, menneskelig udførelse og menneskelig kontrol).
2. Teknisk robusthed og sikkerhed (dvs. modstandsdygtighed over for angreb og sikkerhed, *fall back-plan* og generel sikkerhed, nøjagtighed, pålidelighed og reproducerbarhed).
3. Overholdelse af privatlivets fred og datastyring (dvs. respekt for privatlivets fred, datakvalitet og -integritet samt adgang til data).
4. Gennemsigtighed (dvs. sporbarhed, forklarlighed og kommunikation).
5. Diversitet, ikke-diskrimination og retfærdighed (dvs. undgåelse af urimelig bias, tilgængelighed og universelt design samt inddragelse af interessenter).
6. Samfunds- og miljømæssig velfærd (dvs. bæredygtighed og miljøvenlighed, social indvirkning, samfund og demokrati).
7. Ansvarlighed (dvs. mulighed for revision, minimering og rapportering af negative virkninger, afvejninger og klageadgang). (Europa-Kommissionen, 2019.)

Morley et al. (2019) har konstrueret en typologi ved at koble de etiske principper til faserne i AI-livscyklussen for at sikre, at AI-systemet designes, implementeres og udrulles på en etisk korrekt måde. Typologien viser, at hvert etiske princip bør overvejes i hver fase af AI-livscyklussen. Typologien giver et kort øjebliksbillede af, hvilke værktøjer der i øjeblikket er tilgængelige for AI-udviklere for at fremme udviklingen af etisk AI – fra principper til praksis. Hele typologien fra Morley et al. kan findes på https://tinyurl.com/AppliedAIEthics.

4.2. Kende til forskellige etiske rammer og retningslinjer samt deres anvendelse i virkelige scenarier, hvor AI-systemer indgår

Kende til forskellige etiske rammer og retningslinjer samt deres anvendelse i virkelige scenarier, hvor AI-systemer indgår. Analysere forskellige etiske rammer og retningslinjer, der er relevante for AI og algoritmisk bias. Anvende etiske rammer på virkelige AI-scenarier for at vurdere de etiske dimensioner af AI-systemer. Sammenligne og kontrastere forskellige etiske rammer for at afgøre deres egnethed til at håndtere algoritmisk bias i specifikke AI-applikationer.

Rammer og retningslinjer for AI-etik: Etiske rammer og retningslinjer består af værdier, principper, standarder og handlinger, der skal vejlede enkeltpersoner, grupper, samfund, institutioner og virksomheder i den private sektor for at sikre indlejring af etik i alle faser af AI-systemets livscyklus. De sigter mod at beskytte, fremme og respektere menneskerettigheder og grundlæggende frihedsrettigheder, menneskelig værdighed og lighed; at beskytte nuværende og fremtidige generationers interesser; at bevare miljøet, biodiversiteten og økosystemerne; og at respektere kulturel mangfoldighed i alle faser af AI-systemets livscyklus (European Commission, 2019; UNESCO, 2021).

Forskellige sæt af etiske principper og rammer for AI publiceres typisk af IT-branchen (f.eks. Google, IBM, Microsoft, Intel), regeringer/styrende organer (f.eks. UK Lords Select Committee, Europa-Kommissionens Ekspertgruppe på Højt Niveau om Kunstig Intelligens) og den akademiske verden (f.eks. Future of Life Institute, IEEE, AI4People) (Zhou et al., 2020). Der er ikke ét enkelt etisk princip, der eksplicit bakkes op af alle andre eksisterende etiske rammer og retningslinjer, men der er en begyndende konvergens omkring følgende principper: gennemsigtighed, retfærdighed og rimelighed, ansvar, ikke-skadelighed, privatlivets fred, velgørenhed, frihed og autonomi, tillid, bæredygtighed, værdighed og solidaritet (Jobin et al, 2019).

UNESCO (2021) udgav i november 2021 den første globale standard nogensinde om AI-etik – *Recommendation on the Ethics of Artificial Intelligence*. Denne standard blev vedtaget af alle 193 medlemsstater. Beskyttelsen af menneskerettigheder og værdighed er hjørnestenen i anbefalingerne, som er baseret på fremme af deres grundlæggende principper.

4.3. Forstå vigtigheden af gennemsigtighed, ansvarlighed og inddragelse af interessenter i den etiske AI-udviklingsproces

Forstå vigtigheden af gennemsigtighed, ansvarlighed og inddragelse af interessenter i den etiske AI-udviklingsproces. Designe AI-systemer, der fremmer gennemsigtighed, så brugere og interessenter kan forstå beslutningsprocessen. Implementere ansvarlighedsforanstaltninger for at sikre, at AI-udviklere og organisationer tager ansvar for konsekvenserne af deres AI-systemer. Fremme interessenternes engagement i AI-udvikling, herunder at anmode om input fra forskellige perspektiver og overveje behovene hos forskellige samfundsgrupper, der er berørt af AI-systemer.

Gennemsigtighed: Gennemsigtighed (eller forklarbarhed) er afgørende for at opbygge og fastholde brugernes tillid til AI-systemer. Det betyder, at processer skal være eksplicitte, at AI-systemers evner og formål skal kommunikeres åbent, og at beslutninger – i det omfang det er muligt – skal kunne forklares til dem, der er direkte og indirekte berørt af systemerne. Uden sådanne oplysninger kan en beslutning ikke anfægtes på behørig vis.

Det er ikke altid muligt at forklare, hvorfor en model har genereret et bestemt output eller en bestemt beslutning (og hvilken kombination af inputfaktorer, der har bidraget til det). Disse tilfælde omtales som "black box"-algoritmer og kræver særlig opmærksomhed. Under disse omstændigheder kan det være nødvendigt med andre forklaringsforanstaltninger (f.eks. sporbarhed, revisionsmuligheder og gennemsigtig kommunikation om systemfunktioner), forudsat at systemet som helhed respekterer grundlæggende rettigheder. I hvor høj grad der er behov for forklaringer, afhænger især af konteksten og konsekvensernes alvor, hvis outputtet er fejlagtigt eller på anden måde unøjagtigt (Europa-Kommissionen, 2019).

Ansvarlighed: Ansvarlighed er en af hjørnestenene i styringen af AI. Ansvarlighed kan defineres på mange måder, men i bund og grund er det en forpligtelse til at informere om og begrunde sin adfærd over for en myndighed (Novelli, Taddeo & Floridi, 2023). Generelt set handler ansvarlighed inden for AI om forventningen om, at designere, udviklere og de, der implementerer

systemerne, overholder standarder og lovgivning for at sikre, at AI-systemer fungerer korrekt i løbet af deres livscyklus (Fjeld et al., 2020).

Der bør udvikles passende mekanismer, som kan udføre tilsyn, konsekvensvurderinger, revisioner og *due diligence*, herunder beskyttelse af whistleblowere, for at sikre ansvarlighed i forbindelse med AI-systemer og deres indvirkninger på omverdenen i hele systemernes livscyklus. Både tekniske og institutionelle designs bør sikre, at AI-systemer kan revideres og spores, især for at håndtere eventuelle konflikter med menneskerettighedsnormer og -standarder samt trusler mod miljø og økosystem (UNESCO, 2021).

Inddragelse af interessenter: Det er vigtigt at konsultere relevante interessentgrupper, mens man udvikler og styrer brugen af AI-applikationer (Fjeld et al., 2020). Deltagelse af forskellige interessenter i hele AI-systemets livscyklus er nødvendig for at få nogle inkluderende tilgange til AI-governance, så alle kan få fordele af outputtet, og så der sker en bæredygtig udvikling. Interessenter omfatter, men er ikke begrænset til, regeringer, mellemstatslige organisationer, den tekniske branche, civilsamfundet, forskere og den akademiske verden, medier, uddannelsesinstitutioner, politiske beslutningstagere, private virksomheder, menneskerettigheds-institutioner og ligestillingsorganer, overvågningsorganer for antidiskrimination og grupper for unge og børn (UNESCO, 2021).

Inddragelse af interessenter er med til at sikre, at AI-teknologier udvikles og implementeres på en måde, der stemmer overens med værdier, behov og problemstillinger hos forskellige individer og grupper, der er berørt af teknologien. Men i praksis kan det være svært at identificere de mennesker og organisationer, der påvirkes af udviklingen og brugen af AI-systemer. Disse spørgsmål bliver endnu vigtigere, hvis AI-systemer udvikles i projekter med en slags midlertidige organisationer, hvis teammedlemmer måske ikke er tilgængelige, når projektet afsluttes. Nogle AI-systemer kører helt automatisk og tillader ikke menneskelig indgriben, hvilket gør dem unikke i forhold til andre informationssystemer. Disse systemer kan påvirke menneske-, borger- og arbejdstagerrettigheder; de har karakteristika, der ligger tæt på medicinal- og sundhedsbranchen med hensyn til deres påvirkning af menneskeheden, og de har karakteristika, der ligger tæt på infrastrukturprojekter med hensyn til deres påvirkning af miljøet (Miller, 2022).

CU5 – Casestudier og projekter EQF6 LÆRINGSUDBYTTE		
	Efter endt læringsforløb vil deltageren være i stand til at	
VIDEN	**FÆRDIGHEDER**	**HOLDNINGER**
Teoretisk/faglig viden i: CUK5.1. Forstå algoritmisk bias og dens konsekvenser i den virkelige verden. CUK5.2. Være bevidst om kompleksiteten og nuancerne i algoritmisk bias. CUK5.3. Forstå de strategier og værktøjer, der er til rådighed for at løse problemer med algoritmisk bias i praksis.	Færdighedsmål 1: Anvende viden om algoritmisk bias i casestudier fra den virkelige verden og et hands-on-projekt. 1.1. Undersøge casestudier fra den virkelige verden for at identificere tilfælde af algoritmisk bias og dens konsekvenser. 1.2. Udvikle løsninger til at håndtere algoritmisk bias i casestudier under hensyn til de etiske og praktiske implikationer. 1.3. Designe, implementere og evaluere et hands-on-projekt til at tackle algoritmisk bias på et udvalgt område. Færdighedsmål 2: Analysere algoritmisk bias, dens årsager og konsekvenser. 2.1. Analysere forskellige typer af algoritmiske bias og deres underliggende årsager. 2.2. Vurdere den mulige indvirkning af algoritmisk bias på enkeltpersoner og samfundet som helhed. 2.3. Forstå udfordringerne og begrænsningerne ved at identificere, måle og afbøde effekterne af algoritmiske bias. Færdighedsmål 3: Vurdere og integrere tilgængelige strategier og værktøjer til at håndtere problemer med algoritmisk bias. 3.1. Undersøge og evaluere eksisterende strategier til håndtering af algoritmisk bias, herunder fairness-bevidst ML og *bias-audits*. 3.2. Udvikle færdigheder i at bruge værktøjer og teknikker til at opdage, måle og afbøde effekterne af algoritmisk bias i AI-systemer. 3.3. Sammenkoble tværfaglige tilgange, såsom input fra domæneeksperter og forskellige, andre perspektiver, for at forbedre retfærdigheden i AI-systemer.	Holdningsmål 1: Bestræbe sig på at have en problemløsende tankegang i forhold til algoritmisk bias og dens konsekvenser i den virkelige verden. 1.1. Bestræbe sig på at anvende teoretisk viden i praktiske situationer for at håndtere algoritmisk bias. 1.2. Anerkende vigtigheden af praktisk erfaring for at få en dybere forståelse af algoritmisk bias og dens konsekvenser i den virkelige verden. 1.3. Udvikle en problemløsende tankegang, der søger at skabe fair og retfærdige AI-løsninger. Holdningsmål 2: Have en kritisk og samarbejdsorienteret indstilling til at forstå kompleksiteten og nuancerne i algoritmisk bias. 2.1. Fremme empati for enkeltpersoner og samfundsgrupper, der er berørt af algoritmisk bias. 2.2. Fremme et kritisk mind-set, der forstår kompleksiteten og begrænsningerne ved at adressere algoritmisk bias. 2.3. Forstå vigtigheden af tværfagligt samarbejde og forskellige perspektiver i håndteringen af udfordringerne med algoritmisk bias. Holdningsmål 3: Have en åben og samarbejdsvillig indstilling til at bruge de strategier og værktøjer, der er til rådighed for at løse problemer med algoritmisk bias i praksis. 3.1. Have en proaktiv tilgang til at bruge tilgængelige strategier og værktøjer til at bekæmpe algoritmisk bias. 3.2. Værdsætte vigtigheden af kontinuerlig læring og holde sig informeret om ny udvikling i håndteringen af algoritmisk bias. 3.3. Fremme et samarbejdsmiljø, der opfordrer til åben dialog, videndeling og innovation i håndteringen af problemer med algoritmisk bias.

KOMPETENCE/ER I DIGCOMP2.2., ENTRECOMP OG GREENCOMP FRAMEWORKS, SOM CU5 ER FORBUNDET MED	FÆRDIGHEDER FRA DIGCOMP2.2., ENTRECOMP OG GREENCOMP-RAMMEVÆRKERNE, SOM CU5 ER FORBUNDET MED	HOLDNINGER I DIGCOMP2.2., ENTRECOMP OG GREENCOMP FRAMEWORKS, SOM CU5 ER FORBUNDET MED
- Er opmærksom på, at AI-algoritmer måske ikke kun er konfigureret til at give den information, som brugeren ønsker; de kan også indeholde et kommercielt eller politisk budskab (f.eks. for at opmuntre brugerne til at blive på hjemmesiden, se eller købe noget bestemt, dele bestemte meninger). Dette kan også have negative konsekvenser (f.eks. reproduktion af stereotyper, deling af misinformation). (AI) (DigiComp 2.2.)	- Etisk og bæredygtig tankegang (f.eks. tråde, der opfører sig etisk, være ansvarlig, vurdere indvirkninger): Vurdere konsekvenserne og virkningen af ideer, muligheder og handlinger. (...) Handle ansvarligt. (EntreComp)	- Er åben for AI-systemer, der støtter mennesker i at træffe informerede beslutninger i overensstemmelse med deres mål (f.eks. brugere, der aktivt beslutter, om de vil handle på baggrund af en anbefaling eller ej) (DIGCOMP2.2.)
- Er opmærksom på, at de data, som AI er afhængig af, kan indeholde bias. Hvis det er tilfældet, kan disse bias blive automatiseret og forværret ved brug af AI. For eksempel kan søgeresultater om forskellige erhverv indeholde stereotyper om mande- eller kvindejobs (f.eks. mandlige buschauffører, kvindelige sælgere). (AI) (DigiComp 2.2.)	- Udtryk for bæredygtighedsværdier:	- Være parat til at overveje etiske spørgsmål i forbindelse med AI-systemer (f.eks. i hvilke sammenhænge, såsom domfældelse af kriminelle, bør AI-anbefalinger ikke bruges uden menneskelig indgriben?) (AI) (DigiComp 2.2.)
- Ved, at indsamlet og behandlet data, for eksempel af onlinesystemer, kan bruges til at genkende mønstre (f.eks. repetitioner) i nye data (f.eks. andre billeder, lyde, museklik,	1.2 Understøtte retfærdighed - at understøtte lighed og retfærdighed for nuværende og fremtidige generationer og lære af tidligere generationer med henblik på bæredygtighed.	- Kan identificere både de positive og negative konsekvenser af brugen af alle data (indsamling, kodning og behandling), men især personlige data, af AI-drevne digitale teknologier som apps og onlinetjenester. (DIGCOMP2.2.)
onlineadfærd) for yderligere at optimere og personliggøre onlinetjenester (f.eks. reklamer). (DigiComp 2.2.)	1.3 Tage vare på naturen - at anerkende, at mennesker er en del af naturen, og at respektere andre arters og naturens egne behov for og rettigheder til at genoprette og regenerere sunde og modstandsdygtige økosystemer. (GreenComp)	- Overvejer de etiske konsekvenser af AI-systemer i hele deres livscyklus: de omfatter både miljøpåvirkningen (dvs. de miljømæssige konsekvenser af produktionen af digitale enheder og tjenester) og samfundsmæssige konsekvenser, f.eks. platformisering af arbejde og algoritmisk styring, der kan undertrykke arbejdstagernes privatliv eller rettigheder; brugen af billig arbejdskraft til rubricering af billeder for at træne AI-systemer. (DIGCOMP2.2.)
- Er opmærksom på, at sensorer, der bruges i mange digitale teknologier og applikationer (f.eks. ansigtssporingskameraer, virtuelle assistenter, kropsbårne teknologier, mobiltelefoner, smarte enheder), genererer store mængder data, herunder personlige data, der kan bruges til at træne et AI-system. (AI) (DigiComp 2.2.)		- Er åben for at engagere sig i samarbejdsprocesser for at co-designe og co-skabe nye produkter og tjenester baseret på AI-systemer for at understøtte og øge borgernes deltagelse i samfundet. (AI) (DigiComp 2.2.)

- Er bevidst om, at alt, hvad man deler offentligt online (f.eks. billeder, videoer, lyd), kan bruges til at træne AI-systemer. For eksempel kan kommercielle softwarevirksomheder, der udvikler AI-ansigtsgenkendelsessystemer, bruge personlige billeder, der deles online (f.eks. familiefotografier) til at træne og forbedre softwarens evne til automatisk at genkende disse personer på andre billeder, hvilket måske ikke er ønskeligt (f.eks. Kan det krænke privatlivets fred). (AI) (DigiComp 2.2.) - Forstår, at mens anvendelsen af AI-systemer på mange områder normalt er ukontroversiel (f.eks. AI, der hjælper med at afværge klimaforandringer), kan AI, der direkte interagerer med mennesker og træffer beslutninger om deres liv, ofte være kontroversielle (f.eks. CV-sorteringssoftware til rekrutteringsprocedurer, karaktergivning ved eksamener, der kan afgøre om man bliver optaget på en uddannelse eller ej). (AI) (DigiComp 2.2.) - Er i stand til at identificere nogle eksempler på AI-systemer: produktanbefalinger (f.eks. på online shopping sites), stemmegenkendelse (f.eks. af virtuelle assistenter), billedgenkendelse (f.eks. til at opdage tumorer på røntgenbilleder) og ansigtsgenkendelse (f.eks. i overvågningssystemer). (AI) (DigiComp 2.2.)		

5.1. Forstå algoritmisk bias og dens konsekvenser i den virkelige verden

Forstå algoritmisk bias og dens konsekvenser i den virkelige verden. Anvende viden om algoritmisk bias på casestudier fra den virkelige verden og et hands-on-projekt. Undersøge casestudier fra den virkelige verden for at identificere tilfælde af algoritmisk bias og dens konsekvenser. Udvikle løsninger til at håndtere algoritmisk bias i casestudier under hensyn til de etiske og praktiske implikationer. Designe, implementere og evaluere et praktisk projekt til at tackle algoritmisk bias på et udvalgt område.

Algoritmisk bias: Algoritmisk bias refererer til tilstedeværelsen af uretfærdige eller diskriminerende udfald i de resultater, der produceres af et algoritmisk system. Det sker, når en algoritme konsekvent og systematisk favoriserer eller forfordeler visse individer eller grupper baseret på specifikke karakteristika som race, køn eller socioøkonomisk status. I machine learning er algoritmer baseret på flere datasæt, dvs. træningsdata, der specificerer, hvad de korrekte output er for nogle mennesker eller objekter. Ud fra disse træningsdata lærer maskinen en model, som kan anvendes på andre personer eller objekter og forudsige de mulige output for dem. Maskiner kan dog behandle mennesker og objekter, der befinder sig i samme situation, forskelligt. På grund af denne mangel risikerer nogle algoritmer at replikere og endda forstærke menneskelige bias, især dem, der påvirker beskyttede grupper (Friedman & Nissenbaum, 1996; Lange & Duarte, 2017; Lee, Resnick & Barton, 2019).

Implikationer af algoritmisk bias: Algoritmisk bias kan vise sig på flere måder og med varierende grader af konsekvenser for den involverede gruppe. De følgende eksempler illustrerer både en række årsager og virkninger, der enten utilsigtet behandler grupper forskelligt eller bevidst skaber en ulige påvirkning af dem (Lee, Resnick & Barton, 2019):

- *Bias i online rekrutteringsværktøjer* – Internethandelsvirksomheden Amazon stoppede for nylig brugen af en rekrutteringsalgoritme efter at have opdaget kønsbias. AI-softwaren favoriserede ikke de CV'er, der indeholdt ordet "women's" i teksten, og nedprioriterede også CV'er fra kvinder, der gik på kvindekollegier, hvilket resulterede i kønsbias.

- *Bias i ord-associationer* – Forskere fra Princeton University brugte noget standard AI machine learning-software til at analysere og forbinde 2,2 millioner ord. De fandt ud af, at europæiske navne blev opfattet mere positivt end afroamerikanske navne, og at ordene "kvinde" og "pige" var mere tilbøjelige til at blive associeret med kunst i stedet for videnskab og matematik, som var mest tilbøjelige til at blive associeret med mænd.

- *Bias i online-annoncer* – Latanya Sweeney, Harvard-forsker og tidligere teknologichef hos Federal Trade Commission (FTC), fandt ud af, at online-søgeforespørgsler på afroamerikanske navne var mere tilbøjelige til at returnere annoncer til den pågældende person fra en online-tjeneste, der refererede fra arrestjournaler, sammenlignet med annonceresultaterne for hvide navne.

- *Bias i ansigtsgenkendelsesteknologi* – MIT-forskeren Joy Buolamwini fandt ud af, at de algoritmer, der ligger bag tre kommercielt tilgængelige softwaresystemer til ansigtsgenkendelse, ikke genkendte mørkere hudfarver. Generelt anslås de fleste træningsdatasæt til ansigtsgenkendelse at være mere end 75 procent mandlige og mere end 80 procent hvide personer. Når personen på billedet var en hvid mand, kunne softwaren med 99 procents nøjagtighed identificere personen som mand.

- *Bias i strafferetlige algoritmer* – COMPAS-algoritmen (Correctional Offender Management Profiling for Alternative Sanctions), som bruges af dommere til at forudsige, om tiltalte skal tilbageholdes eller løslades mod kaution, mens de venter på retssagen, viste sig ifølge en rapport fra ProPublica at være biased over for afroamerikanere.

5.2. Være bevidst om kompleksiteten og nuancerne i algoritmisk bias

Være bevidst om kompleksiteten og nuancerne i algoritmisk bias. Analysere algoritmisk bias, dens årsager og konsekvenser. Analysere forskellige typer af algoritmisk bias og deres grundlæggende årsager. Evaluere virkningen af algoritmisk bias. Forstå begrænsningerne i at identificere, måle og afbøde effekterne af algoritmisk bias.

Typer og årsager til algoritmisk bias: Der findes forskellige former for algoritmisk bias. Ifølge Friedman og Nissenbaum (1996) kan forskellige typer og årsager til algoritmiske bias inddeles i tre kategorier:

1. Allerede eksisterende bias har sine rødder i sociale institutioner, praksisser og holdninger. Når computersystemer inkluderer bias, der eksisterer uafhængigt, og som regel før systemet blev skabt, så er systemet et eksempel på allerede eksisterende bias. Allerede eksisterende bias kan komme ind i et system enten ved individers eller institutioners eksplicitte og bevidste indsats, eller implicit og ubevidst, selv på trods af de bedste intentioner.

2. Teknisk bias opstår på grund af tekniske begrænsninger eller tekniske overvejelser. Kilder til teknisk bias kan findes i flere dele af designprocessen, herunder begrænsninger i computerværktøjer såsom hardware, software og perifere enheder; processen med at tilføje social mening til algoritmer, der er udviklet uden for en bestemt kontekst; fejl ved generering af pseudo-tilfældige tal; og forsøget på at gøre menneskelige konstruktioner tilgængelige for computere, når vi kvantificerer det kvalitative, diskretiserer det kontinuerlige eller formaliserer det ikke-formelle.

3. Emergent bias opstår i en brugskontekst med rigtige brugere. Denne type bias opstår typisk et stykke tid efter, at et design er færdigt, som et resultat af ændringer i den samfundsmæssige viden, i befolkningen eller i kulturelle værdier. Brugergrænseflader er sandsynligvis særligt udsatte for *emergent bias*, fordi grænseflader qua deres design søger at afspejle de potentielle brugeres evner, karakter og vaner. Derfor kan et skift i brugskonteksten meget vel skabe uhensigtsmæssigheder for nye brugere. Emergent bias opstår, når der fremkommer ny viden i samfundet, som ikke kan blive – eller ikke bliver – inkorporeret i systemet, eller når den befolkningsgruppe, der reelt bruger systemet, adskiller sig fra den befolkningsgruppe, der var de formodede brugere, dvs. der er et mismatch mellem brugere og systemdesign.

5.3. Forstå de strategier og værktøjer, der er til rådighed for at løse problemer med algoritmisk bias i praksis

Forstå de strategier og værktøjer, der er til rådighed for at løse problemer med algoritmisk bias, herunder fairness-bevidst ML og bias-audits. Gennemgå og integrere strategier og værktøjer til håndtering af algoritmisk bias i praksis. Undersøge og evaluere eksisterende strategier til håndtering af algoritmisk bias, herunder fairness-bevidst maskinlæring og bias-audits. Udvikle færdigheder i at bruge værktøjer og teknikker til at opdage, måle og afbøde algoritmisk bias i AI-systemer. Integrere tværfaglige tilgange, såsom input fra domæneeksperter og forskellige perspektiver, for at forbedre AI-systemers retfærdighed.

Strategier og værktøjer til håndtering af problemer med algoritmisk bias: Håndtering af algoritmisk bias kræver omhyggelige overvejelser i de forskellige faser af algoritmers udvikling. Det indebærer at inkludere forskelligartede og repræsentative træningsdata, at identificere og afbøde bias i algoritmens design og beslutningsproces og at udføre omhyggelig testning og evaluering for at opdage og rette op på bias-relaterede problemer. Der er blevet gjort en del for at minimere diskriminerende resultater ved at udvikle

rammer og retningslinjer for at mindske algoritmisk bias, herunder øget gennemsigtighed og ansvarlighed i algoritmesystemer, lovgivningsmæssige foranstaltninger og vedtagelse af fairness-bevidste ML-teknikker (Lee, Resnick & Barton, 2019).

4.2. CHARLIE WP2 "Etisk AI Micro-Credential" EQF4

Dette dybdegående uddannelsesforløb er designet til at give deltagerne en omfattende forståelse af algoritmiske bias og dens implikationer i nutidens datadrevne samfund. Da algoritmer spiller en væsentlig rolle i vores liv og præger beslutningstagningen på tværs af forskellige sektorer, som f.eks. sundhedsvæsnet, uddannelsessektoren, finansverdenen og regeringer, er det afgørende for fagfolk og akademikere/studerende at udvikle en solid forståelse af potentielle bias inden for disse systemer.

Formålet med dette kursus er at give deltagerne den viden og de færdigheder, der er nødvendige for at identificere, afbøde effekterne af og håndtere algoritmiske bias. Gennem en kombination af teoretisk, grundlæggende viden og forskellige former for praktisk anvendelse vil deltagerne opnå indsigt i de etiske, sociale og tekniske aspekter af algoritmisk bias.

Disse formål bidrager til de generelle mål ved at sætte rammerne for de følgende aktiviteter, så lærere, mentorer og elever kan få en klar forestilling om målene for de forskellige uddannelsesveje, og så mulighederne øges hos videregående læreanstalter samt voksen- og ungdomsuddannelsessteder for at tilbyde kurser, der opfylder samfundets behov, men som også er skræddersyet til deltagernes læringsbehov. Helt konkret er de forventede resultater af "Etisk AI Micro-Credential" EQF4 rettet mod at:

1. Udvikle en grundlæggende forståelse af algoritmisk bias, dens årsager og konsekvenser for individer og samfund.
    - 1.1. Tilegne sig viden om definitionen af, kilderne til og manifestationerne af algoritmisk bias.
    - 1.2. Forstå konsekvenserne af biased algoritmer for samfundet og enkeltpersoner.
2. Fremme bevidsthed om og anvendelse af det etiske princip om ikke-skadelighed i AI-udvikling.
    - 2.1. Lære om de risici og skadelige virkninger, der er forbundet med biased algoritmer.
    - 2.2. Udvikle strategier til at minimere skadelige virkninger og fremme etisk AI-udvikling.

3. Forstå vigtigheden af ansvarlighed i AI-systemer og undersøge relevante juridiske og etiske rammer.
    3.1. Undersøge forskellige interessenters rolle i AI-ansvarlighed.
    3.2. Lære om *best practice* for at sikre ansvarlighed i AI-udvikling.
4. Få viden om begrebet gennemsigtighed i AI-systemer og dets betydning for algoritmisk beslutningstagning.
    4.1. Undersøge metoder og værktøjer til at øge gennemsigtigheden i AI.
    4.2. Forstå udfordringerne og begrænsningerne ved at gøre komplekse algoritmer mere forståelige.
5. Undersøge krydsfeltet mellem AI, menneskerettigheder og retfærdighed og deres indvirkning på etisk AI-udvikling.
    5.1. Undersøge, hvordan biased algoritmer kan påvirke menneskerettigheder, såsom ikke-diskrimination, privatlivets fred og ytringsfrihed.
    5.2. Lære strategier til at sikre retfærdighed og lighed i AI-udvikling og -implementering.
6. Anvende etiske principper i AI-udvikling og -implementering gennem praktiske tilgange og scenarier fra den virkelige verden.
    6.1. Undersøge forskellige etiske rammer og retningslinjer samt deres anvendelse i AI-systemer.
    6.2. Forstå vigtigheden af interessenters engagement, tværfagligt samarbejde og implementering af etiske AI-udviklingsprocesser.

Efter at have gennemført dette kursus vil deltagerne have udviklet en omfattende forståelse af algoritmiske bias, deres potentielle indvirkning på forskellige sektorer og de strategier og værktøjer, der er til rådighed for at håndtere disse bias. Denne viden vil være uvurderlig for fagfolk og akademikere/studerende, der arbejder inden for områder, hvor algoritmer bruges til beslutningstagning, og vil hjælpe med at sikre mere lige og retfærdige resultater i en datadrevet verden.

4.2.1. Indhold i EQF4-kurset "Etisk AI Micro-Credential"

Micro-Credential-kurset er struktureret omkring 6 kompetenceenheder (CU'er), der hver især er designet til at udstyre deltagerne med den viden og de færdigheder, der er nødvendige for at navigere i udfordringerne og mulighederne inden for algoritmisk bias.

CU1 – Hvad er algoritmisk bias? I denne del vil de studerende udforske begrebet algoritmisk bias og dets forskellige udtryk. Emnerne omfatter definitionen af algoritmisk bias, årsagerne og kilderne til bias i algoritmer og de mulige konsekvenser af biased algoritmer for enkeltpersoner og samfundet.

CU2 – Ikke-skadelighed *(non-maleficence)* Denne del fokuserer på det etiske princip om ikke-skadelighed, som understreger vigtigheden af at undgå skadelige virkninger i udviklingen og implementeringen af AI-systemer. Deltagerne vil lære om de mulige risici og skadelige virkninger, der er forbundet med biased algoritmer, samt strategier til at minimere skadelige virkninger og sikre en etisk AI-udvikling.

CU3 – Ansvarlighed I delen om ansvarlighed vil de studerende undersøge vigtigheden af at have klare linjer med hensyn til ansvar og ansvarlighed i udviklingen og brugen af AI-systemer. Emnerne omfatter juridiske og etiske rammer for ansvarlighed, de forskellige interessenters rolle og best practice for at sikre ansvarlighed i AI-udviklingen.

CU4 – Gennemsigtighed Denne del dykker ned i begrebet gennemsigtighed i AI-systemer og undersøger vigtigheden af åbenhed, kommunikation og forklarbarhed i algoritmisk beslutningstagning. De studerende vil lære om metoder og værktøjer til at forbedre gennemsigtigheden i AI, samt de udfordringer og begrænsninger, der er forbundet med at gøre komplekse algoritmer mere forståelige.

CU5 – Menneskerettigheder og retfærdighed I delen om menneskerettigheder og retfærdighed vil de studerende udforske krydsfeltet mellem AI, menneskerettigheder og retfærdighed. De vil undersøge, hvordan biased algoritmer kan påvirke menneskerettighederne, herunder retten til ikke-diskrimination, privatlivets fred og ytringsfrihed. De studerende vil også lære om strategier til at sikre retfærdighed og lighed i udviklingen og implementeringen af AI.

CU6 – AI-etik, en praktisk tilgang Denne del fokuserer på den praktiske anvendelse af etiske principper i AI-udvikling og -implementering. De studerende vil udforske forskellige etiske rammer og retningslinjer og lære, hvordan man anvender dem på virkelige scenarier, hvor AI-systemer indgår. Denne del vil også omhandle vigtigheden af at inddrage interessenter, samarbejde tværfagligt og implementere etiske AI-udviklingsprocesser.

4.2.2. Beskrivelse af målgruppe og EQF4-niveau

Den Europæiske Kvalifikationsramme (EQF) er en fælles referenceramme, der forbinder kvalifikationssystemerne i forskellige europæiske lande, hvilket gør

det lettere for enkeltpersoner at få deres kvalifikationer anerkendt i hele Europa (Europa-Kommissionen, 2023). EQF består af otte referenceniveauer, som hver især er defineret af nogle deskriptorer, der angiver de læringsresultater, færdigheder, kompetencer og den selvstændighed, der forventes på det pågældende niveau. EQF-niveau 4 (EQF4) svarer til kvalifikationer, der er opnået på niveau med en ungdomsuddannelse, ikke-akademisk uddannelse eller erhvervskvalifikationer i det europæiske område for videregående uddannelse (EHEA). EQF4-niveauet kan beskrives på følgende måde (European Commission, 2023; CEDEFOP, 2023):

- Faktuel og teoretisk viden: EQF4-kvalifikationer kræver, at deltagerne besidder faktuel og teoretisk viden i en bredere kontekst inden for deres studie- eller erhvervsområde. Dette vidensniveau er nødvendigt for at forstå vigtige principper, koncepter og processer, der er relevante for deres ekspertiseområde.

- Kognitive færdigheder: På EQF4-niveau skal deltagerne udvikle evnen til at anvende deres viden i praksis, analysere situationer og løse problemer ved hjælp af grundlæggende metoder og værktøjer. Dette omfatter kritisk tænkning, problemløsning og beslutningstagning på et mere grundlæggende niveau.

- Praktiske færdigheder: EQF4-kvalifikationer indebærer udvikling af en række praktiske færdigheder, der sætter deltagerne i stand til at udføre opgaver og gennemføre aktiviteter inden for deres valgte område. Disse færdigheder kan også omfatte tekniske evner, brug af relevante værktøjer og udstyr eller evnen til at udføre bestemte processer og procedurer.

- Selvstændighed og ansvar: Elever på EQF4-niveau forventes at udvise nogen grad af selvstændighed og ansvar i deres arbejde. Det indebærer, at de tager ansvar for egen læring og udvikling, kan arbejde under opsyn eller vejledning og kan træffe beslutninger baseret på fastlagte retningslinjer og procedurer.

- Kommunikation og samarbejde: EQF4-kvalifikationer kræver, at deltagerne udvikler effektive kommunikations- og samarbejdsevner, så de kan udveksle information, argumenter eller forslag med andre. Dette omfatter grundlæggende skriftlige, mundtlige og interpersonelle kommunikationsevner samt evnen til at arbejde effektivt i teams eller grupper.

- Livslang læring: På EQF4-niveau skal deltagerne udvikle evnen til at engagere sig i livslang læring, reflektere over deres egne

læringserfaringer og identificere områder, hvor de kan udvikle sig yderligere. Dette inkluderer evnen til at tilpasse sig nye situationer, teknologier eller professionelle miljøer og til at engagere sig i fortsat professionel udvikling.

De kompetencer, der udvikles i løbet af kurset, vil være i tråd med EQF4-deskriptorerne og fokusere på at give deltagerne en bred vifte af viden, kognitive og praktiske færdigheder, selvstændighed, ansvar, kommunikationsevner, samarbejdsevner og udvise en forpligtelse til livslang læring. Kompetencerne for et EQF4-kursus om algoritmisk bias omfatter:

- Faktuel og teoretisk viden: 1.1. Forstå de grundlæggende begreber inden for algoritmer, data og kunstig intelligens. 1.2. Forstå algoritmernes rolle i forskellige sektorer og de mulige konsekvenser af biased beslutningstagning.

- Kognitive færdigheder: 2.1. Identificere og analysere enkle eksempler på algoritmisk bias og dens konsekvenser i den virkelige verden. 2.2. Anvende kritisk tænkning til at evaluere retfærdighed og etiske overvejelser i forbindelse med AI-systemer i en given kontekst.

- Praktiske færdigheder: 3.1. Bruge grundlæggende værktøjer og metoder til at identificere potentielle bias i datasæt og algoritmiske processer. 3.2. Udforske enkle teknikker til at adressere og afbøde bias i AI-systemer.

- Selvstændighed og ansvar: 4.1. Tage ansvar for egen læring og udvikling inden for algoritmisk bias. 4.2. Overveje de etiske implikationer af algoritmisk beslutningstagning i eget arbejde eller på eget studieområde.

- Kommunikation og samarbejde: 5.1. Kommunikere og diskutere emner effektivt med medstuderende, undervisere og andre interessenter; emner, som er relateret til algoritmisk bias. 5.2. Samarbejde med andre om at undersøge og håndtere algoritmisk bias i gruppeprojekter eller casestudier.

- Livslang læring: 6.1. Reflektere over personlige læringsoplevelser og identificere områder for yderligere udvikling inden for algoritmisk bias. 6.2. Udvise en forpligtelse til at holde sig informeret om fremskridt og nye spørgsmål inden for AI-etik og algoritmisk bias.

Ved at fokusere på disse kompetencer vil et EQF4 micro-credential-kursus baseret på algoritmisk bias give deltagerne et solidt fundament inden for området og udstyre dem med den viden, de færdigheder og den forståelse, der er nødvendig for at navigere i de etiske udfordringer, der er forbundet med biased algoritmer og AI-systemer, i deres fremtidige karriere eller videregående uddannelse.

4.2.3. Generel kompetencematrix for EQF4 micro-credential-kurset

LÆRINGSUDBYTTE		
	Efter endt læringsforløb vil deltageren være i stand til at	
VIDEN	**FÆRDIGHEDER**	**HOLDNINGER**
11. Forstå de grundlæggende begreber inden for algoritmer, kunstig intelligens og machine learning og deres anvendelse i forskellige sektorer. 12. Definere algoritmisk bias og identificere dens potentielle kilder, herunder data, modeldesign og implementering. 13. Forstå konsekvenserne af algoritmisk bias for enkeltpersoner, individuelle samfundsgrupper og samfundet som helhed. 14. Forklare de etiske overvejelser om brugen af AI-systemer, herunder retfærdighed, gennemsigtighed og ansvarlighed. 15. Identificere eksempler fra den virkelige verden og casestudier, der illustrerer virkningen af algoritmisk bias og dens etiske implikationer. 16. Forstå vigtigheden af forskelligartede og repræsentative datasæt i udviklingen af AI-systemer. 17. Undersøge forskellige strategier og teknikker til at afbøde og håndtere algoritmisk bias i AI-systemer. 18. Diskutere den rolle, som lovgivning, forordninger og branchestandarder spiller i forhold til at håndtere algoritmisk bias og fremme etisk AI-udvikling. 19. Udforske udfordringerne og mulighederne med hensyn til at skabe balance mellem fordelene ved AI-systemer og de potentielle	1. Analysere og evaluere algoritmer i forhold til deres mulige bias og etiske implikationer. 2. Anvende kritisk tænkning til at vurdere retfærdighed og rimelighed i AI-systemer i virkelige situationer. 3. Bruge grundlæggende statistiske teknikker til at identificere og måle bias i datasæt. 4. Implementere strategier til at håndtere og afbøde bias i udviklingen og implementeringen af AI-systemer. 5. Kommunikere effektivt om algoritmisk bias, dens konsekvenser og potentielle løsninger til både tekniske og ikke-tekniske målgrupper. 6. Samarbejde effektivt i forskellige teams om at udvikle og vurdere AI-systemer i forhold til etisk og retfærdighed. 7. Anvende relevant lovgivning, relevante forordninger og branchestandarder i forbindelse med udvikling og implementering af AI-systemer. 8. Tilpasse og anvende best practices for etisk AI-udvikling i forskellige sektorer og brancher. 9. Besidde problemløsnings-færdigheder med hensyn til at håndtere virkelige tilfælde af algoritmisk bias og udvikle mulige løsninger. 10. Opdatere viden og færdigheder løbende inden for algoritmisk bias og etisk AI-udvikling for at holde	1. Forstå vigtigheden af etiske overvejelser, retfærdighed og lighed i udviklingen og implementeringen af AI-systemer. 2. Udvise en forpligtelse til livslang læring og løbende forbedring i forståelsen og anvendelsen af etiske AI-principper. 3. Udvise ansvarlighed i forbindelse med udvikling, brug og vurdering af AI-systemer. 4. Udvise en proaktiv holdning til at identificere og adressere potentielle bias og etiske problemer i AI-applikationer. 5. Forstå vigtigheden af tværfagligt samarbejde og åben kommunikation i håndteringen af algoritmisk bias og etisk AI-udvikling. 6. Forstå betydningen af gennemsigtighed, ansvarlighed og inddragelse af interessenter i udviklingen af AI-systemer. 7. Respektere de forskellige perspektiver og erfaringer hos personer, der er berørt af AI-systemer, og overveje deres input i design og implementering af retfærdige og unbiased løsninger. 8. Fremme en inkluderende tilgang til AI-udvikling, der tager hensyn til forskellige interessenters behov og bekymringer, herunder marginaliserede og underrepræsenterede samfundsgrupper. 9. Anerkende begrænsningerne ved AI-systemer og vigtigheden af løbende evaluering og forbedring for at minimere eventuelle skadelige virkninger.

risici, der er forbundet med algoritmisk bias. 20. Forstå vigtigheden af tværfagligt samarbejde i udviklingen og implementeringen af retfærdige og etiske AI-systemer, herunder inddragelse af interessenter med forskellige baggrunde og fra forskellige ekspert-områder.	sig ajour med skiftende tendenser og teknologier.	10. Gå ind for etisk AI-udvikling og fremme bevidstheden om algoritmisk bias og dens konsekvenser i professionelle og sociale sammenhænge.

4.2.4. Kompetencematrix for hver CU i EQF4 micro-credential-kurset

CU1 – Hvad er algoritmisk bias? EQF4		
LÆRINGSUDBYTTE		
	Efter endt læringsforløb vil deltageren være i stand til at	
VIDEN	**FÆRDIGHEDER**	**HOLDNINGER**
Teoretisk/faglig viden i: CUK1.1. Definere algoritmisk bias og forstå dens grundlæggende begreber, herunder de faktorer, der bidrager til biased resultater i AI-systemer. CUK1.2. Identificere forskellige typer af algoritmiske bias, såsom datadrevet bias, modeldrevet bias og menneskedrevet bias, og forstå, hvordan de kan komme til udtryk i AI-applikationer. CUK1.3. Forstå de reelle konsekvenser af algoritmisk bias i forskellige sektorer, herunder de potentielle sociale, etiske og juridiske konsekvenser, der er forbundet med biased AI-systemer.	CUS1.1. Analysere AI-systemer og -applikationer for at opdage mulige bias ved hjælp af deltagerens egen forståelse af de grundlæggende begreber inden for algoritmisk bias. CUS1.2. Kategorisere forskellige typer af algoritmiske bias i eksempler fra den virkelige verden ved at bruge deltagerens egen viden om datadrevne, modeldrevne og menneskedrevne bias. CUS1.3. Vurdere virkningen af algoritmisk bias på forskellige sektorer og interessenter under hensyntagen til de potentielle sociale, etiske og juridiske konsekvenser, der er forbundet med biased AI-systemer.	CUA1.1. Udvise en øget bevidsthed om de mulige konsekvenser af algoritmisk bias og indtage en kritisk holdning, når man beskæftiger sig med AI-systemer og -applikationer. CUA1.2. Acceptere vigtigheden af retfærdighed og lighed i AI-systemer og forstå behovet for forskellige perspektiver og inkluderende designprocesser for at minimere bias. CUA1.3. Udvise en vilje til kontinuerlig læring og til at holde sig informeret om den seneste udvikling inden for AI-etik, algoritmisk bias og ansvarlig AI-praksis med det formål at bidrage positivt til udviklingen og brugen af unbiased AI-systemer.

C1 BESKRIVELSE AF VIDENSUDBYTTE

I EQF4-kurset om algoritmisk bias vil deltagerne få grundlæggende viden på området med fokus på at forstå de grundlæggende begreber, identificere forskellige typer bias og kende til de reelle implikationer af algoritmisk bias i forskellige sektorer. Videnudbyttet for dette kursus omfatter:

CU1.1. Definition af algoritmisk bias: De studerende vil lære at definere algoritmisk bias og forstå dens grundlæggende begreber. De vil undersøge de faktorer, der bidrager til biased resultater i AI-systemer, såsom biased dataindsamlingsmetoder, skæve træningsdata og menneskelig beslutningstagning. Denne grundlæggende viden vil hjælpe de studerende med at forstå, hvordan algoritmisk bias kan opstå, og hvilken indvirkning den kan have på AI-applikationer. Det kunne for eksempel være biased ansigtsgenkendelsessystemer, der fejlidentificerer visse demografiske grupper.

CU1.2. Identificering af forskellige typer af algoritmiske bias: Deltagerne vil blive introduceret til forskellige typer af algoritmiske bias, herunder datadrevet bias, modeldrevet bias og menneskedrevet bias. Gennem eksempler og casestudier vil de lære, hvordan disse bias kan vise sig i AI-applikationer, og de vil forstå vigtigheden af at håndtere dem for at sikre retfærdige og unbiased AI-systemer. For eksempel vil de undersøge, hvordan datadrevet bias kan være resultatet af ikke-repræsentative træningsdata, hvilket fører til biased forudsigelser inden for områder som kreditvurdering eller screening af jobansøgere.

CU1.3. Forståelse af de reelle konsekvenser af algoritmisk bias: De studerende vil undersøge de reelle konsekvenser af algoritmisk bias på tværs af forskellige sektorer, såsom sundhedsvæsenet, finansverdenen og strafferet. De vil undersøge de mulige sociale, etiske og juridiske konsekvenser, der er forbundet med biased AI-systemer, og de vil derved få en forståelse for behovet for at minimere algoritmisk bias for at forhindre negative resultater og fremme retfærdighed og lighed i AI-applikationer. Eksempler herpå kan være, hvordan biased AI-systemer i sundhedsvæsenet kan føre til fejldiagnoser eller utilstrækkelig behandling i visse demografiske grupper, eller hvordan algoritmisk bias i strafferetlige systemer kan resultere i uretfærdig domfældelse eller profilering af enkeltpersoner.

CU2 – Ikke-skadelighed EQF4		
LÆRINGSUDBYTTE		
	Efter endt læringsforløb vil deltageren være i stand til at	
VIDEN	**FÆRDIGHEDER**	**HOLDNINGER**
Teoretisk/faglig viden i: CUK2.1. Introducere princippet om ikke-skadelighed: Undervise de studerende i det grundlæggende begreb *non-maleficence* (ikke-skadelighed) og understrege vigtigheden af at undgå skade, når man skaber og bruger AI-systemer, og hvordan dette bidrager til ansvarlig AI-udvikling. CUK2.2. Undersøge mulige skadelige virkninger af biased AI: Vejlede deltagerne i at kunne genkende de forskellige måder, hvorpå biased AI-systemer kan forårsage skade, såsom diskrimination eller krænkelse af privatlivets fred, og bruge eksempler fra den virkelige verden til at illustrere betydningen af at håndtere algoritmisk bias. CUK2.3. Undersøge strategier til at gøre AI-systemer mindre skadelige: Undervise deltagerne i enkle strategier, der kan gøre AI-systemer mindre skadelige, herunder fremme af retfærdighed, ansvar og gennemsigtighed i AI-udvikling og opfordre til samarbejde med eksperter fra forskellige områder.	CUS2.1. Forklare begrebet ikke-skadelighed (*non-maleficence)* i forbindelse med udvikling af kunstig intelligens. CUS2.2. Identificere mulige skadelige virkninger i AI-systemer og forstå vigtigheden af ikke-skadelighed i ansvarlig AI-udvikling. CUS2.3. Analysere forskellige typer af skader, der kan opstå som følge af biased AI-systemer. CUS 2.4. Vurdere eksempler fra den virkelige verden på biased AI-systemer og deres konsekvenser. CUS2.5. Identificere og anvende strategier til at reducere skadelige virkninger fra AI-systemer, såsom at fremme retfærdighed, ansvarlighed og gennemsigtighed. CUS2.6. Samarbejde effektivt med eksperter fra forskellige områder for at håndtere og afbøde skader forårsaget af biased AI-systemer.	CUA2.1. Fremme en følelse af ansvar for at anvende princippet om ikke-skadelighed i AI-udvikling. CUA2.2. Anerkende vigtigheden af etiske overvejelser i AI-systemer, især for at undgå potentielle skadelige virkninger. CUA2.3. Udvikle empati over for enkeltpersoner og samfundsgrupper, der er berørt af skader forårsaget af biased AI-systemer. CUA2.4. Fremme en forpligtelse til at håndtere og afbøde algoritmisk bias i AI-udvikling for at reducere mulige skadelige virkninger. CUA2.5. Have en proaktiv tilgang til implementering af strategier, der reducerer skadelige virkninger fra AI-systemer. CUA2.6. Værdsætte vigtigheden af tværfagligt samarbejde og forskellige perspektiver i forhold til at håndtere og afbøde skader forårsaget af biased AI-systemer.

BESKRIVELSE AF VIDENSUDBYTTE

De studerende vil få grundlæggende viden om ansvarlighed i AI med fokus på at forstå de grundlæggende begreber, de vil kunne klarlægge AI-udvikleres og -brugeres ansvar for at sikre etiske AI-systemer med minimale skadelige virkninger, og de vil kunne forstå de reelle konsekvenser af implementeringen af mekanismer, der fremmer ansvarlighed i AI-systemer.

Videnudbyttet for dette kursus omfatter:

CU2.1. Princippet om ikke-skadelighed: Deltagerne lærer det grundlæggende princip om ikke-skadelighed at kende, og det understreger vigtigheden af at undgå skadelige virkninger, når man skaber og bruger AI-systemer, og hvordan dette bidrager til ansvarlig AI-udvikling.

CU2.2. Mulige skadelige virkninger af biased AI: Deltagerne vil forstå de forskellige måder, hvorpå biased AI-systemer kan forvolde skader, såsom diskrimination eller krænkelse af privatlivets fred, og eksempler fra den virkelige verden vil blive brugt til at illustrere betydningen af at håndtere algoritmisk bias.

CU2.3. Strategier til at gøre AI-systemer mindre skadelige: Deltagerne vil blive fortrolige med enkle strategier, der kan gøre AI-systemer mindre skadelige, herunder fremme af retfærdighed, ansvar og gennemsigtighed i AI-udvikling og tilskyndelse til samarbejde med eksperter fra forskellige områder.

CU3 – Ansvarlighed EQF4		
LÆRINGSUDBYTTE		
	Efter endt læringsforløb vil deltageren være i stand til at	
VIDEN	**FÆRDIGHEDER**	**HOLDNINGER**
Teoretisk/faglig viden i: CUK3.1. Grundlæggende begreb om ansvarlighed i AI: De studerende vil kunne beskrive det grundlæggende begreb om ansvarlighed i AI med vægt på AI-udvikleres og -brugeres ansvar for at sikre, at AI-systemer fungerer etisk korrekt og med minimale skadelige virkninger. CUK3.2. Ansvarlighedens rolle i håndteringen af algoritmisk bias: De studerende vil kunne identificere, hvordan ansvarlighed spiller en afgørende rolle i at forebygge og afbøde virkningerne af algoritmisk bias, og hvordan det at tage ansvar for AI-systemer kan føre til bedre beslutningstagning og mere retfærdige resultater. CUK3.3. Mekanismer, som sikrer ansvarlighed i AI-systemer: De studerende vil blive fortrolige med simple mekanismer, der kan hjælpe med at sikre ansvarlighed i AI-systemer, såsom retningslinjer, netikette, politik for acceptabel brug *(Acceptable Use Policies, AUP)* samt andre forordninger for AI-udviklere og -brugere.	CUS3.1. Forklare begrebet ansvarlighed i forbindelse med udvikling og brug af AI. CUS3.2. Identificere AI-udvikleres og -brugeres ansvar for at sikre etiske AI-systemer med minimale skadelige virkninger. CUS3.3. Analysere forholdet mellem ansvarlighed og algoritmisk bias og forstå dets betydning i AI-udvikling. CUS3.4. Anvende begrebet ansvarlighed i situationer fra den virkelige verden for at opnå bedre beslutningstagning og mere retfærdige resultater. CUS3,5. Identificere og evaluere mekanismer til at sikre ansvarlighed i AI-systemer, såsom retningslinjer, netikette, politik for acceptabel brug *(Acceptable Use Policies, AUP)* samt andre forordninger for AI-udviklere og -brugere. CUS3.6. Vælge og tilpasse disse mekanismer i AI-udvikling og bruge dem til at fremme ansvarlighed og etisk rigtig adfærd.	CUA3.1. Udvikle ansvarsfølelse med hensyn til etisk AI-udvikling og -brug samt anerkende vigtigheden af ansvarlighed. CUA3.2. Forstå betydningen af ansvarlighed for at minimere skadelige virkninger og fremme etisk korrekt adfærd i AI-systemer. CUA3.3. Værdsætte vigtigheden af ansvarlighed i forhold til algoritmisk bias og dens mulige konsekvenser. CUA3.4. Fremme en forpligtelse til at tage ansvar for AI-systemer for at opnå bedre beslutningstagning og mere retfærdige resultater. CUA3.5. Støtte indførelsen og implementeringen af mekanismer, der fremmer ansvarlighed i AI-systemer. CUA3.6. Forstå den rolle, som retningslinjer, netikette, politik for acceptabel brug *(Acceptable Use Policies, AUP)* samt andre forordninger spiller for at fremme en kultur, hvor ansvarlighed og etisk korrekt adfærd er i centrum i udviklingen og brugen af kunstig intelligens.

BESKRIVELSE AF VIDENSUDBYTTE

De studerende vil få grundlæggende viden om ansvarlighed i AI med fokus på at forstå de grundlæggende begreber, de vil kunne klarlægge AI-udvikleres og -brugeres ansvar for at sikre etiske AI-systemer med minimale skadelige virkninger, og de vil kunne forstå de reelle konsekvenser af implementeringen af mekanismer, der fremmer ansvarlighed i AI-systemer.

Vidensudbyttet for dette kursus omfatter:

CU3.1. Grundlæggende begreb om ansvarlighed i AI: De studerende vil lære om det grundlæggende begreb om ansvarlighed i AI. Ansvarlighed i AI er relateret til forventningen om, at designere, udviklere og de, der implementerer AI, vil overholde standarder og lovgivning for at sikre, at AI-systemer fungerer korrekt i løbet af deres livscyklus (Fjeld et al. 2020). De studerende vil kunne identificere AI-udvikleres og -brugeres ansvar for at sikre etiske AI-systemer med minimale skadelige virkninger. De vil udvikle en ansvarsfølelse med hensyn til etisk AI-udvikling og -brug samt anerkende vigtigheden af ansvarlighed og forstå dens betydning for at minimere skadelige virkninger og fremme etisk korrekt adfærd i AI-systemer.

CU3.2. Ansvarlighedens rolle i håndteringen af algoritmisk bias: De studerende vil kunne identificere årsagerne til, at ansvarlighed spiller en afgørende rolle i at forebygge og afbøde virkningerne af algoritmisk bias. De vil lære at analysere forholdet mellem ansvarlighed og algoritmisk bias og forstå vigtigheden af at tage ansvar for AI-systemer med henblik på bedre beslutningstagning og mere retfærdige resultater.

CU3.3. Mekanismer til at sikre ansvarlighed i AI-systemer. De studerende vil blive fortrolige med simple mekanismer, der kan hjælpe med at sikre ansvarlighed i AI-systemer. De studerende vil forstå vigtigheden af at kunne kommunikere effektivt med interessenter (for eksempel ved at give information om AI-systemets rationale og logik, såsom input, output og processer, der bruges til at træffe beslutninger eller give anbefalinger). Eksempler på disse mekanismer kunne være retningslinjer, netikette, politik for acceptabel brug *(Acceptable Use Policies, AUP)* samt andre forordninger for AI-udviklere og -brugere.

CU4 – Gennemsigtighed EQF4		
LÆRINGSUDBYTTE		
	Efter endt læringsforløb vil deltageren være i stand til at	
VIDEN	**FÆRDIGHEDER**	**HOLDNINGER**
Teoretisk/faglig viden i: CUK4.1. Betydningen af gennemsigtighed i AI-systemer: De studerende vil lære om det grundlæggende begreb gennemsigtighed i AI og dets relevans for at sikre, at AI-systemer er forståelige, forklarlige og tilgængelige for interessenter. CUK4.2. Forholdet mellem gennemsigtighed og algoritmisk bias: De studerende vil forstå sammenhængen mellem gennemsigtighed og algoritmisk bias samt kunne identificere farerne ved uigennemsigtighed, og hvordan øget gennemsigtighed kan hjælpe med at identificere, forhindre og afbøde biased resultater i AI-systemer. CUK4.3. Strategier til at fremme gennemsigtighed i AI-systemer: De studerende vil blive fortrolige med enkle strategier til at fremme gennemsigtighed i AI-systemer, såsom at bruge forståelige modeller, tilvejebringe klar og tydelig dokumentation og formidle beslutningsprocesserne i AI-applikationer.	CUS4.1. Forklare begrebet gennemsigtighed og dets betydning i forbindelse med AI-systemer. CUS4.2. Klarlægge fordelene ved transparente AI-systemer for interessenter, herunder forståelighed, forklarlighed og tilgængelighed. CUS4.3. Analysere forholdet mellem gennemsigtighed og algoritmisk bias i AI-systemer. CUS4.4. Analysere farerne ved uigennemsigtighed i AI-systemer og identificere, hvordan begrebet gennemsigtighed kan hjælpe med at forhindre og afbøde biased resultater i AI-systemer. CUS4.5. Identificere og evaluere strategier til fremme af gennemsigtighed i AI-systemer, herunder forståelige modeller, klar dokumentation og effektiv formidling af beslutningsprocesser. CUS4.6. Anvende disse strategier i AI-udvikling og bruge dem til at øge gennemsigtigheden og afbøde effekterne af algoritmisk bias.	CUA4.1. Forstå vigtigheden af gennemsigtighed i AI-systemer for at opbygge tillid og bane vej for en større forståelse hos interessenter. CUA4.2. Forstå gennemsigtighedens rolle i at fremme etisk AI-udvikling og -brug. CUA4.3. Kende til farerne ved uigennemsigtighed i AI-systemer og betydningen af gennemsigtighed i forhold til at håndtere og afbøde effekterne af algoritmisk bias. CUA4.4. Udvikle en forpligtelse til at øge gennemsigtigheden i AI-systemer for at minimere biased resultater. CUA4.5. Acceptere vedtagelsen og implementeringen af strategier, der fremmer gennemsigtighed i AI-systemer. CUA4.6. Forstå den rolle, som forståelige modeller, klar og tydelig dokumentation og effektiv kommunikation spiller for at fremme en kultur af gennemsigtighed og mindske algoritmisk bias.

BESKRIVELSE AF VIDENSUDBYTTE

De studerende vil få viden om betydningen af gennemsigtighed i AI-systemer med fokus på at forstå de grundlæggende begreber, forholdet mellem gennemsigtighed og algoritmisk bias og relevansen af strategierne for at sikre, at AI-systemer er forståelige, forklarlige og tilgængelige for interessenter, idet de også forstår konsekvenserne i den virkelige verden og er klar over, hvor vigtige forståelige modeller, klar dokumentation og effektiv formidling kan være for at fremme en kultur af gennemsigtighed, som mindsker algoritmisk bias.

Vidensudbyttet for dette kursus omfatter:

CU4.1. Betydningen af gennemsigtighed i AI-systemer: De studerende vil kunne definere det grundlæggende begreb gennemsigtighed i AI og dets relevans for at sikre, at AI-systemer er forståelige, forklarlige og tilgængelige for interessenter. De vil kunne definere fordelene ved og forstå vigtigheden af transparente AI-systemer for at opbygge tillid og bane vej for en større forståelse hos interessenter. Et eksempel: En AI-model, der er designet til at opdage kræft, kan blive livstruende for et menneske, selv hvis modellen kun tager 1% fejl. I sådanne tilfælde er AI og mennesker nødt til at arbejde sammen, og opgaven bliver meget lettere, når AI-modellen kan forklare, hvordan den nåede frem til en bestemt beslutning. Gennemsigtighed gør AI til en holdspiller.

CU4.2. Forholdet mellem gennemsigtighed og algoritmisk bias: De studerende vil kunne se sammenhængen mellem gennemsigtighed og algoritmisk bias, samt forstå farerne ved uigennemsigtighed, og hvordan øget gennemsigtighed kan hjælpe med at identificere, forhindre og afbøde biased resultater i AI-systemer. De vil forstå betydningen af gennemsigtighed i forhold til at håndtere og afbøde effekterne af algoritmisk bias. Et eksempel: Ofte er AI-algoritmer uigennemsigtige i den forstand, at alle interessenter ikke har adgang til diverse bagvedliggende forklaringer. Der kan være forskellige kilder til denne uigennemsigtighed. Nogle gange undlader institutioner eller virksomheder at kommunikere, når de er afhængige af AI-systemer, eller hvordan disse systemer fungerer.

CU4.3. Strategier til at fremme gennemsigtighed i AI-systemer. De studerende vil blive fortrolige med enkle strategier til at fremme gennemsigtighed i AI-systemer, såsom at bruge forståelige modeller, tilvejebringe klar og tydelig dokumentation og formidle beslutningsprocesserne i AI-applikationer. De vil forstå, hvor vigtige disse strategier er for at fremme en kultur af gennemsigtighed og mindske algoritmisk bias. Et eksempel: AI kan påvirke forskellige interessenter, såsom brugere, kunder, medarbejdere, ledere, lovgivere eller samfundet. For at sikre gennemsigtighed og ansvarlighed er man nødt til at forpligte og styrke AI-interessenter i hele IS-livscyklussen.

CU5 – Menneskerettigheder og retfærdighed EQF4		
LÆRINGSUDBYTTE		
	Efter endt gamification-læringsforløb vil deltageren være i stand til at	
VIDEN	**FÆRDIGHEDER**	**HOLDNINGER**
Teoretisk/faglig viden i: CUK5.1. Relevansen af menneskerettigheder og retfærdighed i AI-systemer: De studerende lærer vigtigheden af menneskerettigheder og retfærdighed i udviklingen og implementeringen af AI-systemer, og de vil kunne identificere, hvordan dette bidrager til retfærdige resultater og forhindrer skadelige virkninger for enkeltpersoner og forskellige samfundsgrupper. CUK5.2. Krydsfeltet mellem algoritmisk bias og menneskerettigheder: De studerende vil kunne se forbindelsen mellem algoritmisk bias og menneskerettigheder samt identificere, hvordan biased AI-systemer kan påvirke sårbare befolkningsgrupper uforholdsmæssigt meget og bibeholde eksisterende uligheder. CUK5.3. Principper for retfærdighed i AI-systemer: De studerende vil forstå de grundlæggende principper for retfærdighed i AI-systemer, såsom lige muligheder, ikke-diskrimination, proceduremæssig retfærdighed, lighed og retfærdighed og forstå deres rolle med hensyn til at fremme retfærdige resultater for fremtidige generationer.	CUS5.1. Forklare betydningen af menneskerettigheder og retfærdighed i relation til AI-systemer og hvilken rolle disse spiller i at fremme retfærdige resultater og forhindre skadelige virkninger. CUS5.2. Anvende principperne for menneskerettigheder og retfærdighed i forbindelse med udvikling og implementering af AI. CUS5.3. Analysere forholdet mellem algoritmisk bias og menneskerettigheder og anerkende den mulige indvirkning på sårbare befolkningsgrupper, og hvordan uligheder kan bibeholdes. CUS5.4. Identificere eksempler fra den virkelige verden på biased AI-systemer, der har negativ indflydelse på menneskerettighederne, og udtænke strategier til at løse disse problemer. CUS5.5. Definere og skelne mellem forskellige principper for retfærdighed i AI-systemer, herunder lige muligheder, ikke-diskrimination, proceduremæssig retfærdighed, lighed og retfærdighed. CUS5.6. Anvende principperne for retfærdighed i AI-udvikling og bruge dem til at fremme retfærdige resultater og afbøde algoritmisk bias for fremtidige generationer.	CUA5.1. Forstå vigtigheden af menneskerettigheder og retfærdighed i AI-systemer for at fremme lighed og forhindre skadelige virkninger. CUA5.2. Forpligte sig til at overholde menneskerettigheder og opretholde retfærdighed i AI-udvikling og -implementering. CUA5.3. Forstå betydningen af krydsfeltet mellem algoritmisk bias og menneskerettigheder i udformningen af AI-systemers indvirkning på enkeltpersoner og forskellige samfundsgrupper. CUA5.4. Udvikle en følelse af ansvar med hensyn til at håndtere og afbøde virkningerne af biased AI-systemer på menneskerettigheder og sårbare befolkningsgrupper. CUA5.5. Forstå betydningen af forskellige principper for fairness, lighed og retfærdighed i AI-systemer for at afbøde algoritmisk bias. CUA5.6. Fremme en forpligtelse til at indarbejde retfærdighedsprincipper i udvikling og brug af AI for at opnå mere retfærdige resultater, hvor fremtidige generationers interesser tilgodeses.

BESKRIVELSE AF VIDENSUDBYTTE

De studerende vil få viden om vigtigheden af menneskerettigheder og retfærdighed i AI-systemer med fokus på at forstå de grundlæggende begreber, krydsfeltet mellem algoritmisk bias og menneskerettigheder samt principperne for retfærdighed i AI-systemer, idet de også kan se de reelle konsekvenser og kan forstå vigtigheden af forskellige principper for fairness,

lighed og retfærdighed i AI-systemer for at afbøde virkningerne af algoritmisk bias og opnå mere retfærdige resultater, hvor fremtidige generationers interesser tilgodeses.

Vidensudbyttet for dette kursus omfatter:

5.1. Relevansen af menneskerettigheder og retfærdighed i AI-systemer: De studerende vil lære vigtigheden af menneskerettigheder og retfærdighed i udviklingen og implementeringen af AI-systemer for at fremme retfærdige resultater og forhindre skadelige virkninger. De studerende vil forstå, at AI giver vidtrækkende fordele for den menneskelige udvikling, men indebærer også risici (som f.eks. et endnu større skel mellem de privilegerede og de knapt så privilegerede, gradvis kompromittering af individuelle friheder på grund af overvågning, samt at uafhængig tænkning og dømmekraft erstattes af automatiseret kontrol).

5.2. Forstå krydsfeltet mellem menneskerettigheder og algoritmisk retfærdighed: I denne grundlæggende del vil de studerende blive introduceret til det vigtige krydsfelt mellem menneskerettigheder og algoritmisk retfærdighed. Med udgangspunkt i Latours teorier (2005), som understregede teknologiens moralske og etiske konsekvenser, vil de studerende udforske algoritmiske processers rolle og indvirkning på menneskerettighederne. De studerende vil lære om udviklingen af menneskerettigheder i historisk perspektiv, og hvordan denne udvikling relaterer sig til fremkomsten af algoritmiske teknologier. Teoretiske rammer: Introduktion til de teoretiske rammer, hvori samspillet mellem menneskerettigheder og algoritmisk retfærdighed analyseres. Et eksempel: analyse af casestudier, hvor der har været konflikter mellem algoritmiske processer og menneskerettigheder samt debatterne om dem.

5.3. Principper for retfærdighed i AI-systemer: De studerende vil blive fortrolige med de grundlæggende principper for retfærdighed i AI-systemer, såsom lige muligheder, ikke-diskrimination, proceduremæssig retfærdighed, lighed og retfærdighed og anerkende princippernes vigtige status i forhold til at fremme retfærdige resultater for fremtidige generationer.

CU6 – AI-etik, en praktisk tilgang EQF4		
LÆRINGSUDBYTTE		
	Efter endt gamification-læringsforløb vil deltageren være i stand til at	
VIDEN	**FÆRDIGHEDER**	**HOLDNINGER**
Teoretisk/faglig viden i: CUK6.1. Betydningen af AI-etik i praksis: De studerende lærer betydningen af at indarbejde etiske overvejelser i udviklingen og implementeringen af AI-systemer, idet de forstår, hvordan etik kan afbøde de negative konsekvenser af algoritmisk bias og fremme tilliden til AI-applikationer. CUK6.2. Etiske rammer og retningslinjer for AI-systemer: De studerende vil blive fortrolige med forskellige etiske rammer og retningslinjer, der gælder for AI-systemer, såsom AI-etiske principper, ansvarlig AI-praksis og branchespecifikke regler, og forstå deres rolle med hensyn til at styre etisk AI-udvikling. CUK6.3. Anvendelse af AI-etik i den virkelige verden: De studerende vil udforske praktiske eksempler på og casestudier af AI-systemer inden for forskellige områder, analysere de etiske udfordringer og overvejelser i hvert scenarie samt identificere best practice for at sikre etisk AI-udvikling og -implementering.	CUS6.1. Forklare vigtigheden af etik i forbindelse med udvikling og implementering af AI, herunder etikkens rolle med hensyn til at afbøde algoritmisk bias og fremme tillid. CUS6.2. Anlægge etiske betragtninger i forbindelse med udvikling og brug af AI for at sikre ansvarlige og troværdige AI-systemer. CUS6.3. Identificere og analysere forskellige etiske rammer og retningslinjer for AI-systemer, herunder AI-etiske principper, ansvarlig AI-praksis og branchespecifikke regler. CUS 6.4. Anvende etiske rammer og retningslinjer i AI-udvikling og bruge dem til at styre en ansvarlig og etisk AI-praksis. CUS6.5. Analysere eksempler fra den virkelige verden og casestudier af AI-systemer for at identificere etiske udfordringer og overvejelser inden for forskellige områder. CUS6.6. Anvende best practices og etiske retningslinjer på virkelige scenarier, hvor AI-systemer er involveret, for at sikre etisk udvikling og implementering.	CUA6.1. Anerkende vigtigheden af etik i AI-udvikling og -implementering for at mindske algoritmisk bias og fremme tillid. CUA6.2. Forpligte sig til at indarbejde etiske overvejelser i AI-systemer for at sikre ansvarlige og troværdige AI-applikationer. CUA6.3. Forstå den rolle, som etiske rammer og retningslinjer spiller for etisk udvikling og brug af AI. CUA6.4. Udvikle en forpligtelse til at anvende etiske rammer og retningslinjer for at sikre en ansvarlig og etisk AI-praksis. CUA6.5. Anerkende betydningen af at forstå og håndtere etiske udfordringer i den virkelige verdens AI-applikationer. CUA6.6. Fremme en forpligtelse til at anvende best practice og etiske retningslinjer i virkelige scenarier for at sikre etisk AI-udvikling og -implementering.

BESKRIVELSE AF VIDENDSUDBYTTE

Målsætning: Denne del er udtænkt således, at studerende på EQF4-niveau kan dykke ned i de praktiske aspekter af AI-etik og indarbejde principper for ansvarlig AI-udvikling og -brug i virkelige scenarier. De studerende opfordres til at være proaktive med hensyn til implementeringen af etiske retningslinjer i AI-miljøer ved at trække på deres teoretiske viden og omsætte den i praksis.

6.1. Forstå de etiske implikationer af AI

I dette indledende modul vil de studerende indledningsvis få et solidt grundlag med hensyn til de forskellige etiske dimensioner inden for AI. De studerende vil blive introduceret til de mulige konsekvenser af AI for samfundet,

enkeltpersoner og hele verden set i lyset af Floridi et al.'s (2018) forståelse af de etiske udfordringer vedrørende AI.

Forklaring:

Etisk grundlag: Analyse og forståelse af de forskellige etiske teorier og deres implikationer for AI.

Moralske implikationer: Detaljeret undersøgelse af de moralske implikationer, der stammer fra AI og andre automatiserede beslutningstagningssystemer.

Eksempel: Analyse af casestudier fra den virkelige verden, der skildrer de etiske dilemmaer, man står over for i AI-branchen.

6.2. Identifikation og reduktion af uetisk praksis i AI

Denne del handler om at håndtere behovet for at kunne se og reducere en mulig uetisk praksis inden for AI. Med udgangspunkt i Bostroms (2014) overvejelser om fremtiden for AI og hvordan den stemmer overens med menneskelige værdier, lærer de studerende, hvordan man identificerer og forhindrer uetisk praksis i udvikling og anvendelse af kunstig intelligens.

Forklaring:

Identifikation af uetisk praksis: En undersøgelse af forskellige tegn på uetisk praksis i AI.

Strategier til afhjælpning: Udvikling af strategier til at reducere potentielle uetiske resultater, der stammer fra AI-applikationer.

Eksempel: Rollespilsaktiviteter, hvor deltagerne identificerer og håndterer uetisk praksis i simulerede AI-miljøer.

6.3. Praktisk anvendelse af etiske retningslinjer inden for AI

Studerende i dette modul bliver hjulpet til at omsætte teoretisk viden til praktisk handling. De studerende vil fokusere på aktiviteter, der fremmer praktisk anvendelse af etiske retningslinjer i AI-scenarier, set i lyset af Ryan & Stahls (2020) forståelse heraf, som netop fokuserer på udformningen af etiske retningslinjer.

Forklaring:

Udvikling af etiske retningslinjer: Udarbejdelse af etiske retningslinjer, der kan anvendes i AI-projekter i den virkelige verden.

Implementering i virkelige scenarier: Workshops og aktiviteter med fokus på at implementere de udviklede retningslinjer i virkelige AI-projekter.

Eksempel: De studerende skal deltage i et projekt, hvor de udvikler og anvender etiske retningslinjer i et virkeligt AI-projekt.

6.4. Fremme ansvarlig AI-udvikling og -implementering

Den sidste del vil fokusere på at fremme et miljø, der tilskynder til ansvarlig AI-udvikling og -implementering. De studerende vil udforske metoder til at fremme globalt samarbejde og skabelsen af ansvarlig AI, som er baseret på Jobin et al.'s foreslåede rammer (2019), der giver et globalt perspektiv på AI-etik.

Forklaring:

Globalt samarbejde: Forståelse af vigtigheden af globalt samarbejde for at fremme ansvarlig AI.

Samfundsengagement: Opfordring til samfundsengagement og deltagelse i etisk AI-udvikling.

Eksempel: De studerende vil udvikle strategier og planer for at fremme samfundsengagement og globalt samarbejde inden for AI-etik.

4.3. CHARLIE WP2 læringsudbytte for voksne og unge EQF2 (SERIOUS GAME)

"Unravelling Algorithmic Bias", som kan oversættes til "Bagom Algoritmisk Bias", er et undervisende og interaktivt *serious game* (dvs. et avanceret læringsspil), der introducerer deltagere på EQF2-niveau til det vigtige emne: algoritmisk bias. I løbet af spillet udforskes grundlæggende begreber om algoritmer, deres potentielle bias og de konsekvenser, som disse bias kan have på forskellige dele af samfundet.
Det primære formål med spillet "Bagom Algoritmisk Bias" er at skabe et engagerende læringsmiljø, hvor deltagerne kan undersøge emnet gennem en række interaktive aktiviteter og scenarier. Ved at præsentere indholdet i et tilgængeligt og underholdende format forventes det, at deltagerne får en dybere forståelse og bevidsthed om betydningen af algoritmisk bias.

Det forventede udbytte af denne WP er helt konkret at:

- Introducere deltagerne til begrebet algoritmisk bias og dets relevans i dag samt fremhæve vigtigheden af at håndtere dette emne i forskellige sektorer.
- Præsentere deltagerne for virkelige eksempler på algoritmisk bias og illustrere de mulige konsekvenser og etiske implikationer, der er forbundet med biased algoritmer.
- Fremme deltagernes evne til kritisk tænkning, så de kan analysere og evaluere algoritmer og deres potentielle bias i forskellige sammenhænge.
- Udvikle deltagernes forståelse af forskellige typer af algoritmisk bias, herunder datadrevet, modeldrevet og menneskedrevet bias.
- Opfordre deltagerne til at udforske betydningen af retfærdighed, gennemsigtighed og ansvarlighed i udviklingen og implementeringen af AI-systemer.
- Øge deltagernes bevidsthed om de etiske principper og retningslinjer, der styrer AI-udvikling, såsom ikke-skadelighed, menneskerettighederne og privatlivets fred.
- Udfordre deltagerne til at anvende deres nyvundne viden til at identificere potentielle bias i simulerede AI-applikationer og foreslå løsninger til at afbøde effekterne af disse bias.
- Fremme samarbejde og kommunikation mellem deltagerne, når de arbejder sammen om at løse algoritmisk bias-relaterede problemer og dele deres viden.
- Engagere deltagerne i aktiviteter, som de kan reflektere over, og opfordre dem til at overveje deres egne perspektiver og bias, og hvordan disse kan påvirke brugen af AI-systemer.
- Inspirere deltagerne til at blive proaktive fortalere for etisk AI-udvikling og -anvendelse i deres lokalsamfund og på deres arbejdssteder, hvilket fremmer en følelse af ansvar og samfundsengagement.

Når spillet er færdigt, vil deltagere på EQF2-niveau have fået en grundlæggende forståelse for algoritmisk bias og de mulige konsekvenser, dette kan have. De vil udvikle grundlæggende færdigheder i problemløsning, der gør dem i stand til at genkende bias i AI-applikationer. Derudover vil deltagerne have forbedret deres kommunikations- og samarbejdsevner, samtidig med at de vil opnå en større ansvarsfølelse og større bevidsthed om etiske overvejelser i brugen af AI og teknologi.

4.3.1. Indhold i spillet "Bagom Algoritmisk Bias"

Strukturen af spillets hovedindhold for deltagere på EQF2-niveau kan inddeles i fem moduler:

1. Introduktion til algoritmisk bias:
Grundlæggende begreber inden for algoritmer og AI
Introduktion til bias og fairness i AI-systemer
Eksempler på algoritmisk bias fra den virkelige verden

2. Typer af algoritmiske bias:
Data-drevet bias
Model-drevet bias
Menneske-drevet bias
Eksempler og scenarier for hver type bias

3. Identificering og håndtering af algoritmisk bias:
Strategier til at opdage bias i AI-applikationer
Grundlæggende teknikker til at afbøde effekterne af bias
Enkeltpersoners og organisationers rolle i håndteringen af algoritmisk bias

4. Etiske overvejelser i AI:
Betydningen af etik i udvikling og brug af AI
Introduktion til etiske principper som ikke-skadelighed, ansvarlighed og gennemsigtighed
Sammenhængen mellem menneskerettigheder, retfærdighed og AI-etik

5. Spilaktiviteter og udfordringer:
Interaktive scenarier, hvor spillerne identificerer og håndterer bias i AI-applikationer
Teambaseret problemløsning relateret til etisk AI-udvikling
Refleksions- og diskussionssessioner for at styrke læringen og fremme en bevidsthed om algoritmisk bias

I løbet af spillet vil deltagerne deltage i samarbejdsaktiviteter og udfordringer, der er designet til at styrke deres forståelse af begreberne og samtidig fremme teamwork, kommunikation og problemløsningsevner.

4.3.2. Beskrivelse af målgruppe og EQF2-niveau

Den Europæiske Kvalifikationsramme (EQF) niveau 2 (EQF2) udgør et grundlæggende kompetenceniveau inden for forskellige emner eller

færdigheder (Europa-Kommissionen, 2023). Deltagere på EQF2-niveau er typisk på de tidlige stadier af deres læringsrejse eller i gang med en grundlæggende erhvervsuddannelse. EQF2-niveauet kan i hovedtræk beskrives på følgende måde:

- Grundlæggende viden: EQF2-deltagere har en grundlæggende forståelse af nøglebegreber, principper og idéer inden for deres valgte fagområde, dog uden at kunne gå i dybden med komplekse teorier eller avancerede emner.
- Praktiske færdigheder: Deltagere på EQF2-niveau kan anvende deres grundlæggende viden til at udføre enkle opgaver og aktiviteter relateret til deres fagområde ved hjælp af grundlæggende værktøjer og teknikker under opsyn eller med vejledning.
- Kognitive færdigheder: EQF2-elever kan anvende grundlæggende kognitive færdigheder, såsom problemløsning og beslutningstagning, i velkendte og ukomplicerede situationer. De kan følge enkle instruktioner og procedurer for at fuldføre deres opgaver.
- Kommunikation: EQF2-deltagere er i stand til at viderebringe grundlæggende information og idéer på en klar og præcis måde, både mundtligt og skriftligt, i sammenhænge, som er velkendte for dem.
- Samarbejde og teamwork: På EQF2-niveau kan deltagerne arbejde effektivt sammen med andre i et team eller en gruppe, idet de demonstrerer samarbejdsvillighed og en forståelse af grundlæggende interpersonelle dynamikker.
- Ansvar og selvstændighed: EQF2-elever kan tage ansvar for deres handlinger og beslutninger inden for rammerne af deres begrænsede viden og færdigheder. De kan arbejde under opsyn, følge instruktioner og søge hjælp, når det er nødvendigt.
- Tilpasningsevne og fleksibilitet: Deltagere på EQF2-niveau kan tilpasse sig enkle ændringer i deres lærings- eller arbejdsmiljø og kan håndtere grundlæggende udfordringer eller nederlag med vejledning og støtte.
- Selvbevidsthed og selvrefleksion: EQF2-elever kan reflektere over deres læringsoplevelser og identificere områder, hvor de kan forbedre sig, hvilket viser et grundlæggende niveau af selvbevidsthed og en vilje til personlig vækst.
- Etisk bevidsthed: På EQF2-niveau kan deltagerne forstå de grundlæggende etiske principper og retningslinjer, der er relevante for deres fagområde, og udvise en forståelse af deres betydning i hverdagssituationer.
- Livslang læring: EQF2-elever opfordres til at udvikle en interesse for deres valgte fagområde og være indstillet på livslang læring, hvilket

fremmer nysgerrighed og en vilje til at fortsætte deres uddannelse og kompetenceudvikling ud over EQF2-niveauet.

4.3.3. Generel kompetencematrix for spillet "Bagom Algoritmisk Bias" på EQF2-niveau

LÆRINGSUDBYTTE		
	Efter afslutning af spillet vil deltageren være i stand til at	
VIDEN	**FÆRDIGHEDER**	**HOLDNINGER**
1. Forstå det grundlæggende koncept for algoritmer, og hvordan de bruges i hverdagen, herunder enkle eksempler på deres anvendelse i forskellige sektorer. 2. Forstå vigtigheden af retfærdighed og unbiased beslutningstagning i forbindelse med algoritmiske processer. 3. Forstå det grundlæggende begreb om algoritmisk bias og have en forståelse af, hvordan bias kan opstå i AI-systemer og hvilke mulige konsekvenser det kan have. 4. Identificere enkle eksempler på algoritmisk bias i virkelige situationer, hvilket tydeliggør vigtigheden af at håndtere disse bias for at opnå retfærdige resultater. 5. Forstå den rolle, som data spiller i AI-systemer, og hvordan biased eller ufuldstændige data kan føre til biased beslutninger. 6. Få en grundlæggende bevidsthed om etiske overvejelser inden for AI og vigtigheden af at udvikle og implementere unbiased algoritmer. 7. Forstå behovet for gennemsigtighed og ansvarlighed i AI-systemer for at sikre ansvarlig brug heraf samt håndtere potentielle bias. 8. Forstå betydningen af menneskerettigheder og retfærdighed i forbindelse med AI og den rolle, som unbiased algoritmer spiller i forhold til at fremme disse principper.	1. Analysere enkle scenarier fra den virkelige verden for at identificere tilfælde af algoritmisk bias og dens mulige konsekvenser. 2. Anvende grundlæggende kritisk tænkning til at vurdere retfærdigheden i AI-systemer og systemernes beslutningsprocesser. 3. Udvikle enkle strategier til at minimere forekomsten af bias i AI-systemer, f.eks. ved at bruge mere forskelligartede og repræsentative data. 4. Kommunikere effektivt om grundlæggende begreber inden for algoritmisk bias og konsekvenserne heraf til fagfæller og ikke-specialister. 5. Bruge problemløsningsevner til at løse grundlæggende udfordringer relateret til algoritmisk bias og retfærdighed i AI-systemer. 6. Reflektere over personlige erfaringer og fordomme for bedre at forstå vigtigheden af at fremme retfærdighed og unbiased beslutningstagning i AI-systemer. 7. Samarbejde med andre om at identificere mulige problemer relateret til algoritmisk bias og udvikle passende løsninger. 8. Udvise grundlæggende etisk bevidsthed i forbindelse med AI og forstå vigtigheden af ansvarlig AI-udvikling og -implementering. 9. Tilpasse sig forskellige situationer, der involverer algoritmisk bias, og se behovet for kontinuerlig læring og fremskridt. 10. Udvise en proaktiv holdning til at håndtere algoritmisk bias, tage	1. Forståelse af vigtigheden af retfærdighed og unbiased beslutningstagning i AI-systemer, hvor man kan se de mulige konsekvenser af algoritmisk bias for enkeltpersoner og samfundet. 2. Etisk ansvarlighed, hvor man erkender den enkeltes rolle i udviklingen og implementeringen af AI-systemer og behovet for at handle etisk og ansvarligt. 3. Åbenhed over for forskellige perspektiver, hvor man er villig til at overveje forskellige synspunkter og erfaringer, når man behandler spørgsmål relateret til algoritmisk bias og retfærdighed. 4. Forpligtelse til kontinuerlig læring og personlig vækst, forståelse for, at håndtering af algoritmisk bias kræver løbende læring og tilegnelse. 5. Samarbejde og teamwork, hvor man værdsætter andres bidrag og anerkender vigtigheden af at arbejde sammen om at løse komplekse problemer såsom algoritmisk bias. 6. Kritisk tænkning, sætte spørgsmålstegn ved antagelser og være villig til at revurdere personlige overbevisninger og fordomme i lyset af nye oplysninger eller perspektiver. 7. Empati og medfølelse, hvor man overvejer, hvordan algoritmisk bias påvirker forskellige individer og samfund, og hvor man stræber efter at fremme retfærdighed og lighed i AI-systemer. 8. Respekt for privatlivets fred og databeskyttelse, hvor man forstår vigtigheden af at opretholde brugernes privatliv og beskytte personlige oplysninger i AI-applikationer.

9. Udvikle en grundlæggende bevidsthed om de potentielle juridiske og sociale konsekvenser af algoritmisk bias og behovet for passende regler og retningslinjer. 10. Udvikle en interesse for at lære mere om algoritmisk bias, AI-etik og relaterede emner, hvilket fremmer en forpligtelse til livslang læring og ansvarligt engagement i AI-teknologier.	initiativ til at lære mere og engagere sig i relevante diskussioner og aktiviteter.	9. Social bevidsthed, forståelse af de større, samfundsmæssige konsekvenser af algoritmisk bias og behovet for fælles handling for at imødegå disse udfordringer. 10. Proaktivt engagement, tage initiativ til at deltage i diskussioner og aktiviteter relateret til algoritmisk bias og retfærdighed, og være fortaler for ansvarlig AI-udvikling og -implementering.

Referencer

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. ProPublica, May, 23.

Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Abstraction in Sociotechnical Systems. In ACM Conference on Fairness, Accountability, and Transparency (pp. 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. Available 20.6.2023 at 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. Harvard University. Available at 20.6.2023 https://dash.harvard.edu/bitstream/handle/1/42160420/HLS%20White%20Paper%20Final_v3.pdf?sequence=1&isAllowed=y

Floridi, L., Cowls, 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., Cowls, 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. Available 20.6.2023 at https://nissenbaum.tech.cornell.edu/papers/biasincomputers.pdf

Jobin, A., Ienca, M. & Vayena, E. (2019). Artificial Intelligence: The Global Landscape of Ethics Guidelines. Available at 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. Available at 20.6.2023 https://medium.com/impact-engineered/understanding-bias-in-algorithmic-design-db9847103b6e

Latour, B. (2005). Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford University Press.

Lee, N. T., Resnick, P., & Barton, G. (2019). Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms. Available 20.6.2023 at 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, December 2022. Available at 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. Available at 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, Etchmendy 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. Available 20.6.2023 at 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. Available 20.6.2023 at 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.