



Etisk AI-mikrocertifikat

HÆFTE

**CU5 | Menneskerettigheder og
retfærdighed**

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

Hvordan bruger man denne flipbook?

Dette dokument er interaktivt. I hele dokumentet finder du links til yderligere information.



Knap, der fører dig til begyndelsen af dokumentet. Dette ikon vises i øverste højre hjørne af siderne.



Når du ser denne pil, betyder det, at du har en **interaktiv farvetekst** at klikke på, som er forbundet med et eksternt link.

ANSVARSRASKRIVELSE: Bemærk, at vi ikke kan garantere den fortsatte tilgængelighed af eksternt indhold, f.eks. videoer, da de kan ændres eller fjernes af deres forfattere eller værtsplatforme.

Indeks

Klik på menuen

01. Introduktion

02. Relevansen af menneskerettigheder og retfærdighed i AI-systemer

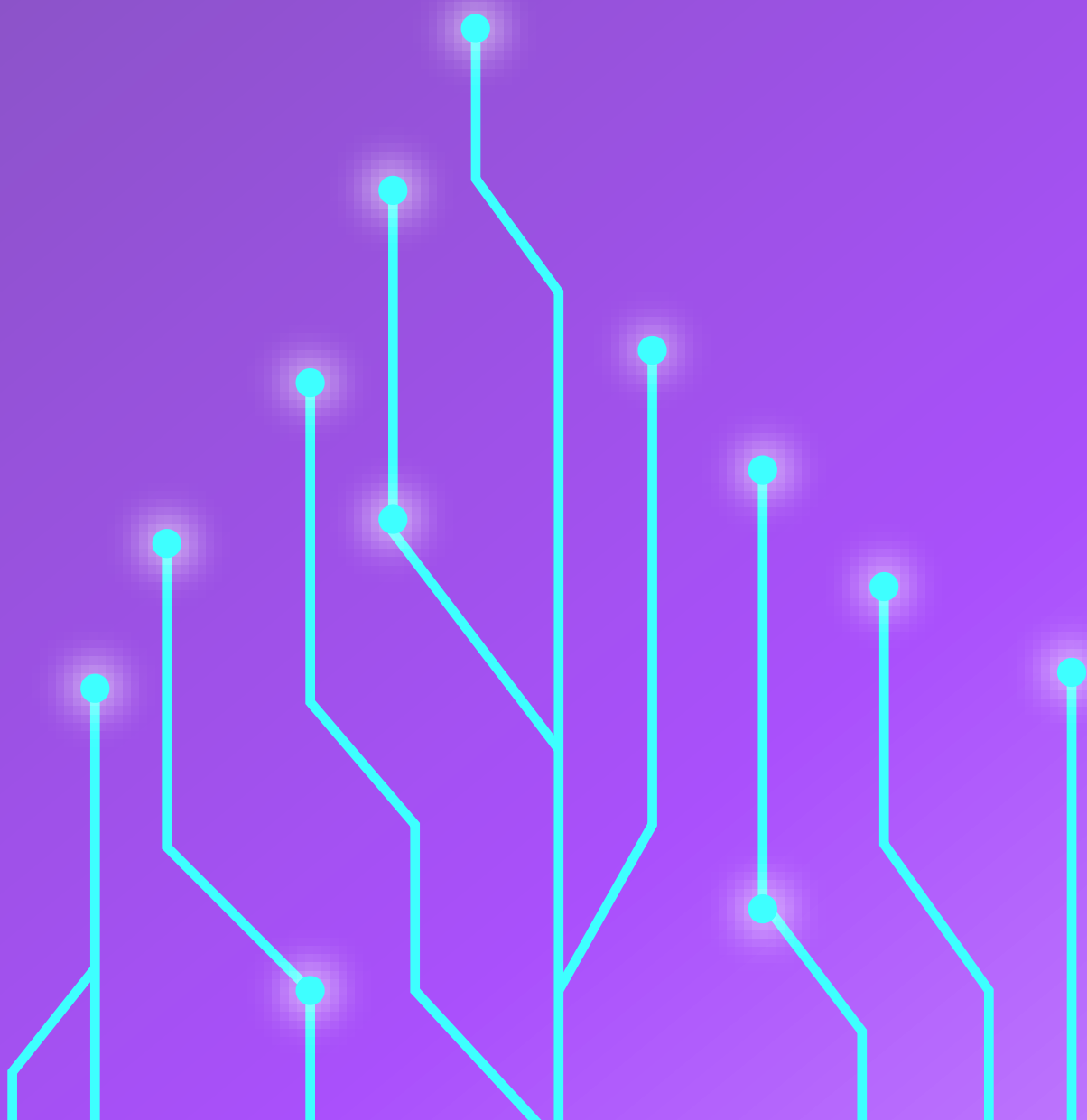
03. Skæringspunktet mellem menneskerettigheder og algoritmisk retfærdighed

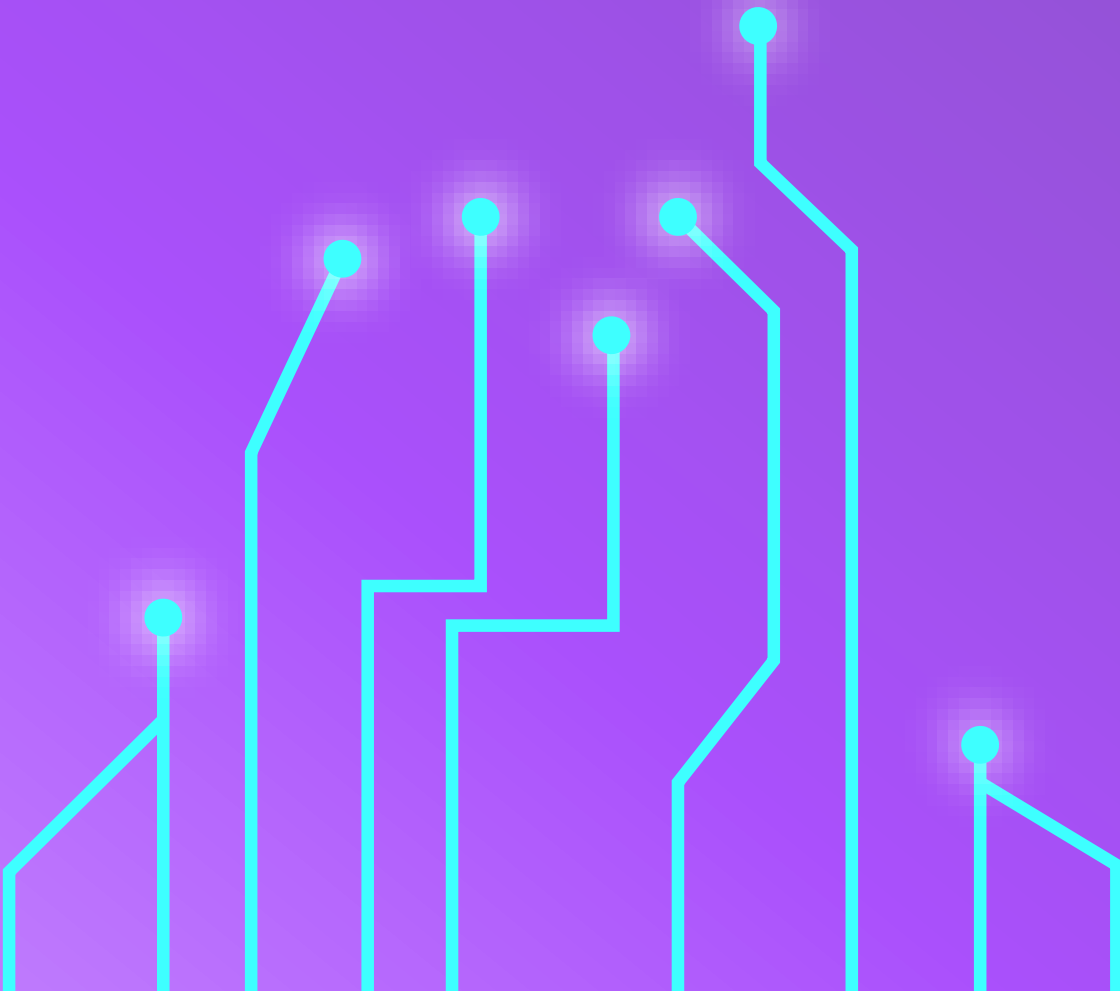
04. Principper for retfærdighed i AI-systemer

05. Konklusion

01. Introduktion

CU5 | Menneskerettigheder og retfærdighed





01. Introduktion

I denne kompetenceenhed vil de studerende tilegne sig omfattende viden om den afgørende rolle, som menneskerettigheder og retfærdighed spiller inden for AI-systemer. Denne udforskning vil dykke ned i grundlæggende koncepter, der understreger skæringspunktet mellem algoritmisk bias og menneskerettigheder samt principperne om retfærdighed, der er iboende i AI-systemer. Ved at forstå disse begreber vil eleverne sætte pris på de virkelige konsekvenser og værdien af retfærdighedsprincipper, lighed og retfærdighed til at afbøde algoritmisk bias og fremme mere retfærdige resultater med behørig hensyntagen til fremtidige generationers interesser.

Resultaterne af dette kursus omfatter:

- **Relevansen af menneskerettigheder og retfærdighed i AI-systemer** for at fremme retfærdige resultater og forebygge skade inden for udvikling og anvendelse af AI. Vi vil præsentere de potentielle fordele og risici, der er forbundet med AI-teknologier, og anerkende behovet for at tage fat på spørgsmål som socioøkonomisk ulighed, krænkelse af privatlivets fred og udhuling af autonomi.



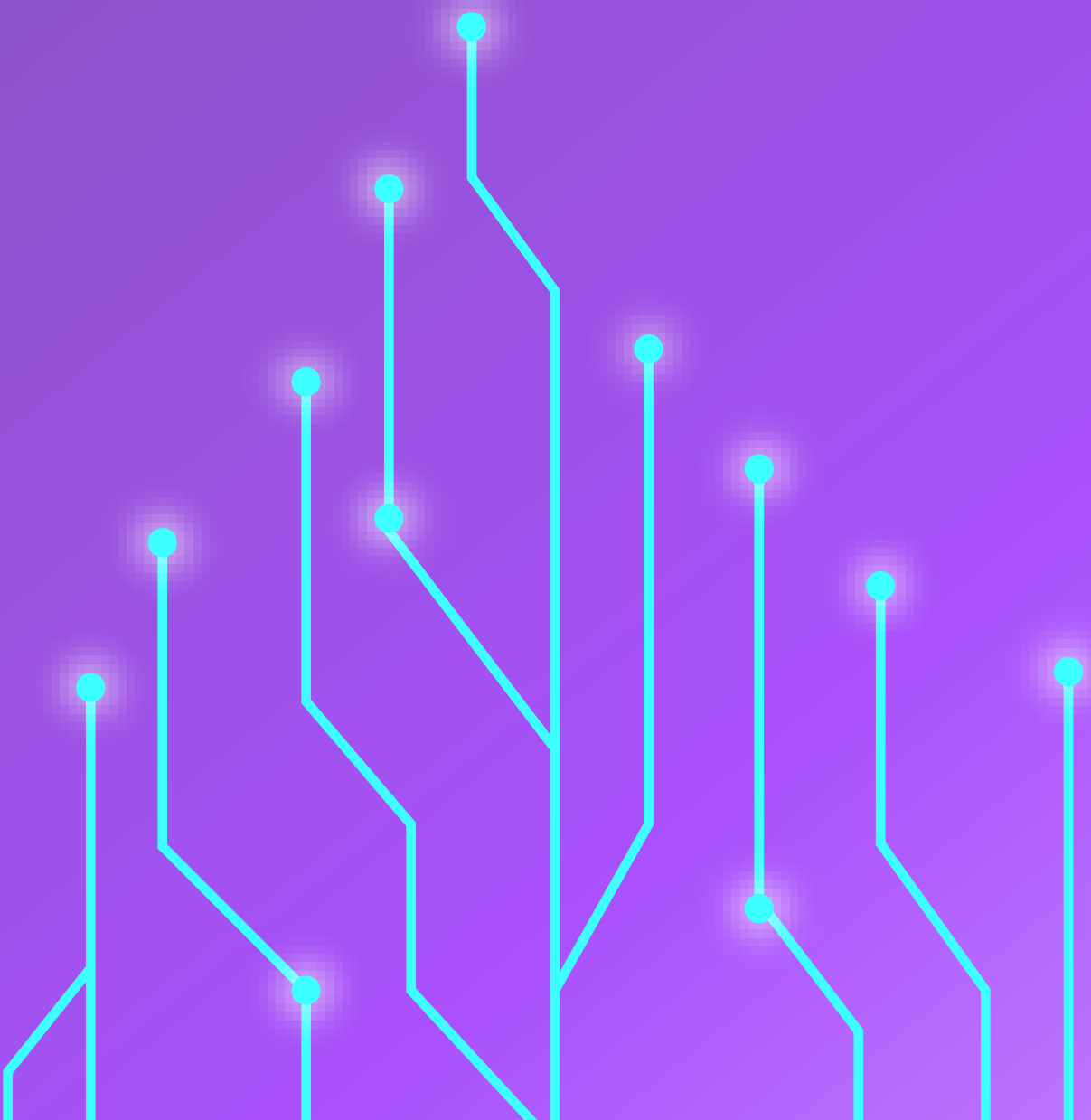


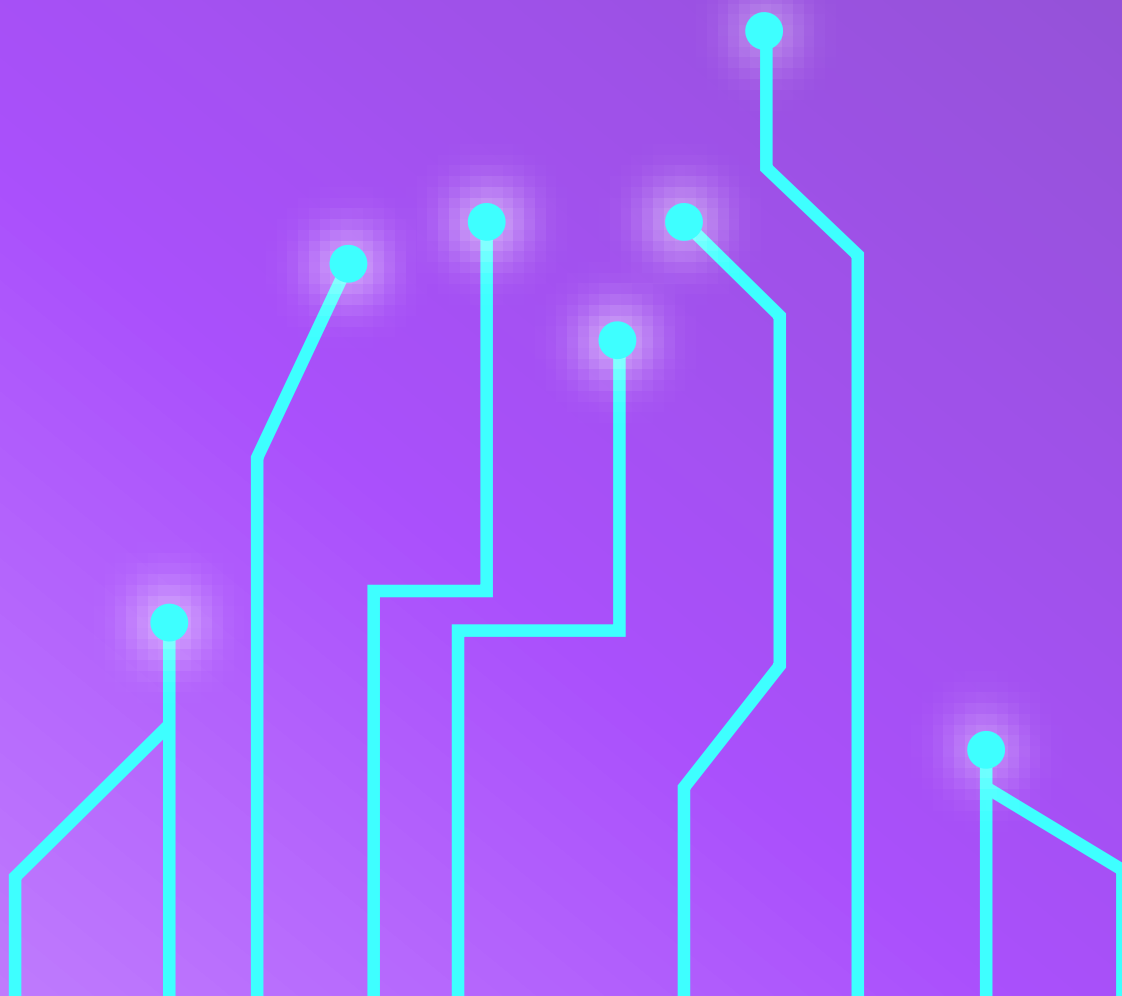
- **Krydsfeltet mellem menneskerettigheder og algoritmisk retfærdighed:** Denne grundlæggende enhed vil introducere de studerende til det kritiske krydsfelt mellem menneskerettigheder og algoritmisk retfærdighed, ved at trække på teoretiske rammer for at analysere de etiske konsekvenser af teknologi. Ved at studere den historiske udvikling og nutidige debatter vil eleverne forstå, hvordan algoritmiske processer påvirker menneskerettighederne, og udforske casestudier, der illustrerer konflikter mellem algoritmiske teknologier og menneskerettigheder.
- **Principper for retfærdighed i AI-systemer er,** lige muligheder, ikke-diskrimination, proceduremæssig retfærdighed, lighed og retfærdighed. De studerende skal anerkende vigtigheden af disse principper i udformningen af retfærdige resultater for fremtidige generationer, og de vil blive rustet til at navigere i etiske udfordringer og advokere for ansvarlig AI-udvikling og -implementering.

I de senere år har integrationen af systemer med kunstig intelligens (AI) i forskellige aspekter af samfundet, givet anledning til betydelige etiske overvejelser om beskyttelse af menneskerettigheder og fremme retfærdighed. Denne enhed har til formål at dykke ned i den kritiske betydning af menneskerettigheder og retfærdighed i AI-systemer og deres centrale rolle i at fremme retfærdige resultater og samtidig forhindre skade.

02. Relevansen af menneskerettigheder og retfærdighed i AI-systemer

CU5 | Menneskerettigheder og retfærdighed





02. Relevansen af menneskerettigheder og retfærdighed i AI-systemer

Betydningen af menneskerettigheder og retfærdighed i AI-systemer kan ikke overvurderes.

Menneskerettighederne, som er nedfældet i forskellige internationale erklæringer og konventioner, fungerer som hjørnestenen i etisk styring og samfundsmæssig velfærd. I forbindelse med AI-systemer sikrer anvendelsen af menneskerettighedsprincipper, at individers værdighed, autonomi og privatliv respekteres og beskyttes. Desuden er retfærdighed i AI-systemer afgørende for at fremme retfærdige resultater og mindske opretholdelsen af eksisterende fordomme og uligheder i samfundet.

AI-systemer har potentiale til at påvirke forskellige aspekter af menneskelivet, lige fra beskæftigelsesmuligheder og adgang til tjenester til retspleje og beskyttelse af borgerlige frihedsrettigheder. Derfor er det afgørende at beskytte menneskerettighederne og fremme retfærdighed i udviklingen og implementeringen af AI for at forhindre skade og opretholde etiske standarder. Ved at integrere menneskerettigheder og retfærdighedsprincipper i AI-systemer kan udviklere mindske risikoen for diskriminerende praksis, krænkelse af privatlivets fred og uretfærdige resultater og derved fremme tillid og ansvarlighed blandt interessenter.



Anvendelse af principperne om menneskerettigheder og retfærdighed i forbindelse med udvikling og implementering af AI kræver en tværfaglig tilgang, der tager højde for juridiske, etiske og samfundsmæssige konsekvenser. Udviklere skal overholde etablerede menneskerettighedsrammer, såsom Verdenserklæringen om Menneskerettigheder og den internationale konvention om borgerlige og politiske rettigheder, for at sikre, at AI-systemer opretholder grundlæggende rettigheder og friheder.

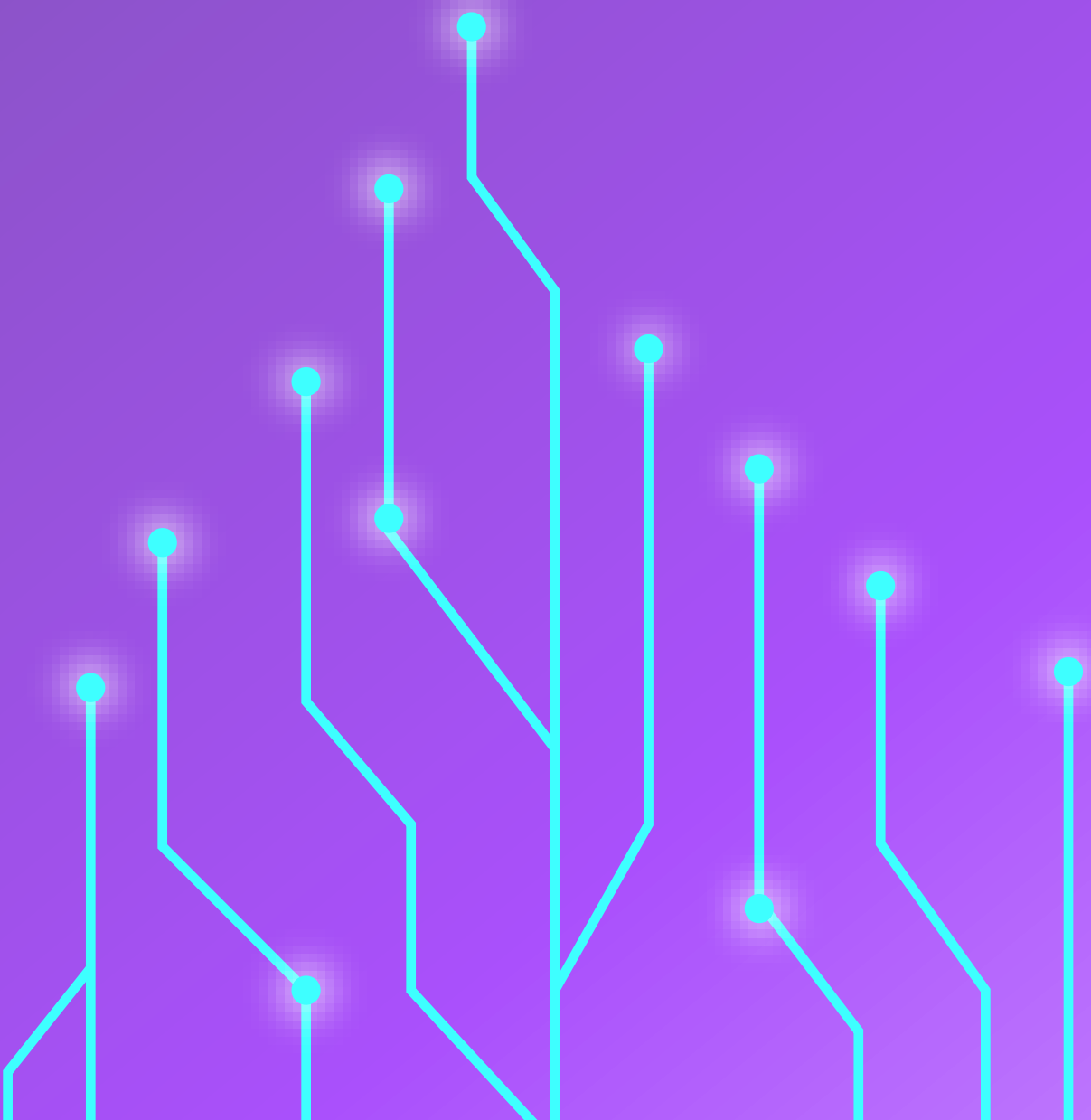
Et vigtigt aspekt ved at anvende menneskerettigheder og retfærdighedsprincipper i AI-udvikling er at sikre gennemsigtighed og ansvarlighed i hele AI-livscyklussen. Gennemsigtighed gør det muligt for interessenter at forstå, hvordan AI-systemer fungerer, hvilke data de bruger, og hvilke beslutningsprocesser der er involveret. Ved at levere klar dokumentation og forklaringer kan udviklere give brugerne mulighed for at vurdere AI-systemernes retfærdighed og pålidelighed og holde dem, der er ansvarlige for deres udvikling og implementering, ansvarlige.

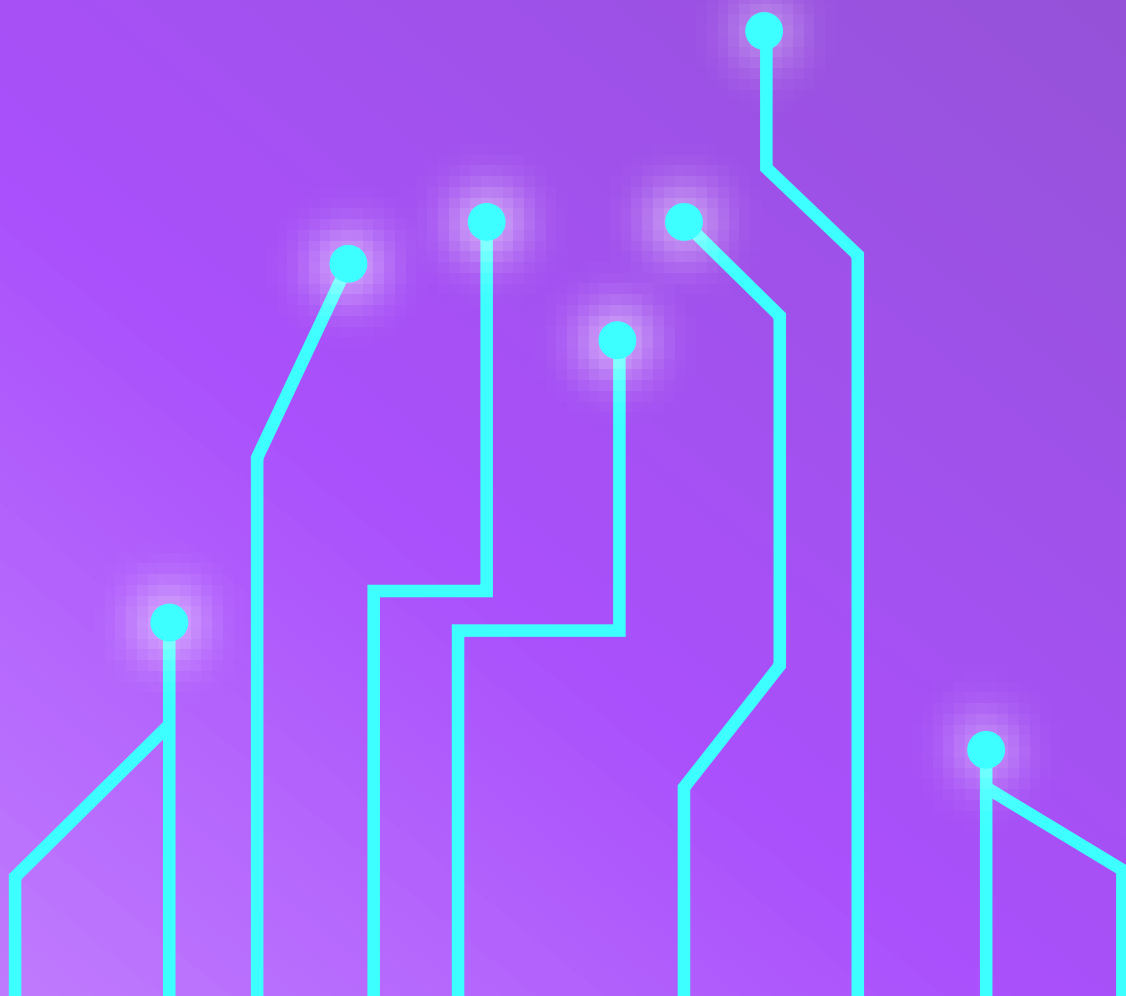
Desuden kræver retfærdighed i AI-udvikling nøje overvejelse af strategier til afhjælpning af bias og algoritmiske ansvarlighedsmekanismer. Udviklere skal proaktivt identificere og adressere bias i træningsdata, algoritmedesign og beslutningsprocesser for at forhindre diskriminerende resultater. Derudover sikrer implementering af mekanismer til tilsyn og erstatning, at personer, der er berørt af AI-systemer, har mulighed for at klage i tilfælde af uretfærdig behandling eller skade.



03. Krydsfeltet mellem menneskerettigheder og algoritmisk retfærdighed

CU5 | Menneskerettigheder og retfærdighed





03. Krydsfeltet mellem menneskerettigheder og algoritmisk retfærdighed

Forholdet mellem algoritmisk bias og menneskerettigheder er et spørgsmål med mange facetter.

Forholdet mellem algoritmisk bias og menneskerettigheder er et komplekst og mangesidet emne, der kræver omhyggelig undersøgelse. I denne enhed vil de studerende dykke ned i disse forviklinger og anerkende den potentielle indvirkning på sårbare befolkninger og opretholdelsen af uligheder. Ved at forstå den dynamik, der er på spil, kan de studerende identificere eksempler fra den virkelige verden på forudindtagede AI-systemer, der påvirker menneskerettighederne, og udtænke strategier til at løse disse problemer effektivt.

Algoritmisk bias refererer til de systematiske og uretfærdige præferencer eller fordomme, der kan være til stede i data, algoritmer eller beslutningsprocesser i AI-systemer. Når algoritmisk bias ikke kontrolleres, kan det have store konsekvenser for menneskerettighederne, især for sårbare befolkningsgrupper som racemindretal, kvinder, ældre og personer med handicap.

Ved at fastholde eksisterende uligheder og forstærke diskriminerende praksis **kan** forudindtagede **AI-systemer underminere grundlæggende menneskerettighedsprincipper som lighed, ikke-diskrimination og retten til privatliv.**



Indvirkningen af algoritmisk bias på menneskerettighederne kan vise sig på forskellige måder i forskellige sektorer og sammenhænge. For eksempel kan forudindtagede algoritmer, der bruges i ansættelsesprocesser, resultere i diskriminerende praksis, der nægter enkeltpersoner lige muligheder for ansættelse baseret på irrelevante faktorer som race, køn eller alder. På samme måde kan forudseende politialgoritmer i det strafferetlige system være uforholdsmæssigt rettet mod marginaliserede samfund, hvilket fører til uretmæssige anholdelser og krænkelser af retten til en retfærdig rettergang.

Desuden kan algoritmisk bias forværre eksisterende forskelle i adgangen til vigtige tjenester som sundhedspleje, bolig og uddannelse. Biased AI-systemer, der bruges til kreditvurdering eller lånegodkendelsesprocesser, kan systematisk forfordele visse demografiske grupper, fastholde økonomiske uligheder og hindre individers evne til at få adgang til økonomiske ressourcer og muligheder for socioøkonomisk fremgang.

Anerkendelse af den potentielle indvirkning af algoritmisk bias på menneskerettighederne er afgørende for at beskytte alle individers rettigheder og værdighed, især dem, der tilhører marginaliserede eller sårbare samfund. Ved at analysere forholdet mellem algoritmisk bias og menneskerettigheder får de studerende en dybere forståelse af de etiske konsekvenser af AI-teknologier og behovet for proaktive foranstaltninger til at håndtere bias og fremme retfærdighed og lighed.

➤ **Eksempler fra den virkelige verden på forudindtagede AI-systemer, der påvirker menneskerettighederne, og strategier til at løse disse problemer**

Der er masser af eksempler fra den virkelige verden på forudindtagede AI-systemer, der påvirker menneskerettighederne, hvilket understreger det presserende behov for handling for at løse disse problemer. **Fra diskriminerende ansigtsgenkendelsessystemer til forudindtagede algoritmer, der bruges til strafudmåling, har forudindtagede AI-systemer potentialet til at krænke individers rettigheder og fastholde systemiske uligheder.**

For eksempel har det i forbindelse med retshåndhævelse vist sig, at forudindtagede politialgoritmer i uforholdsmæssig grad retter sig mod minoritetsgrupper, hvilket fører til øget overvågning, uretmæssige anholdelser og krænkelser af enkeltpersoners ret til privatliv og frihed fra vilkårlig tilbageholdelse.

På samme måde kan forudindtagede algoritmer, der bruges til medicinsk diagnose og behandlingsplanlægning, resultere i fejldiagnoser eller utilstrækkelig pleje for visse demografiske grupper, hvilket forværrer sundhedsforskelle og underminerer individers ret til sundhed og velvære.





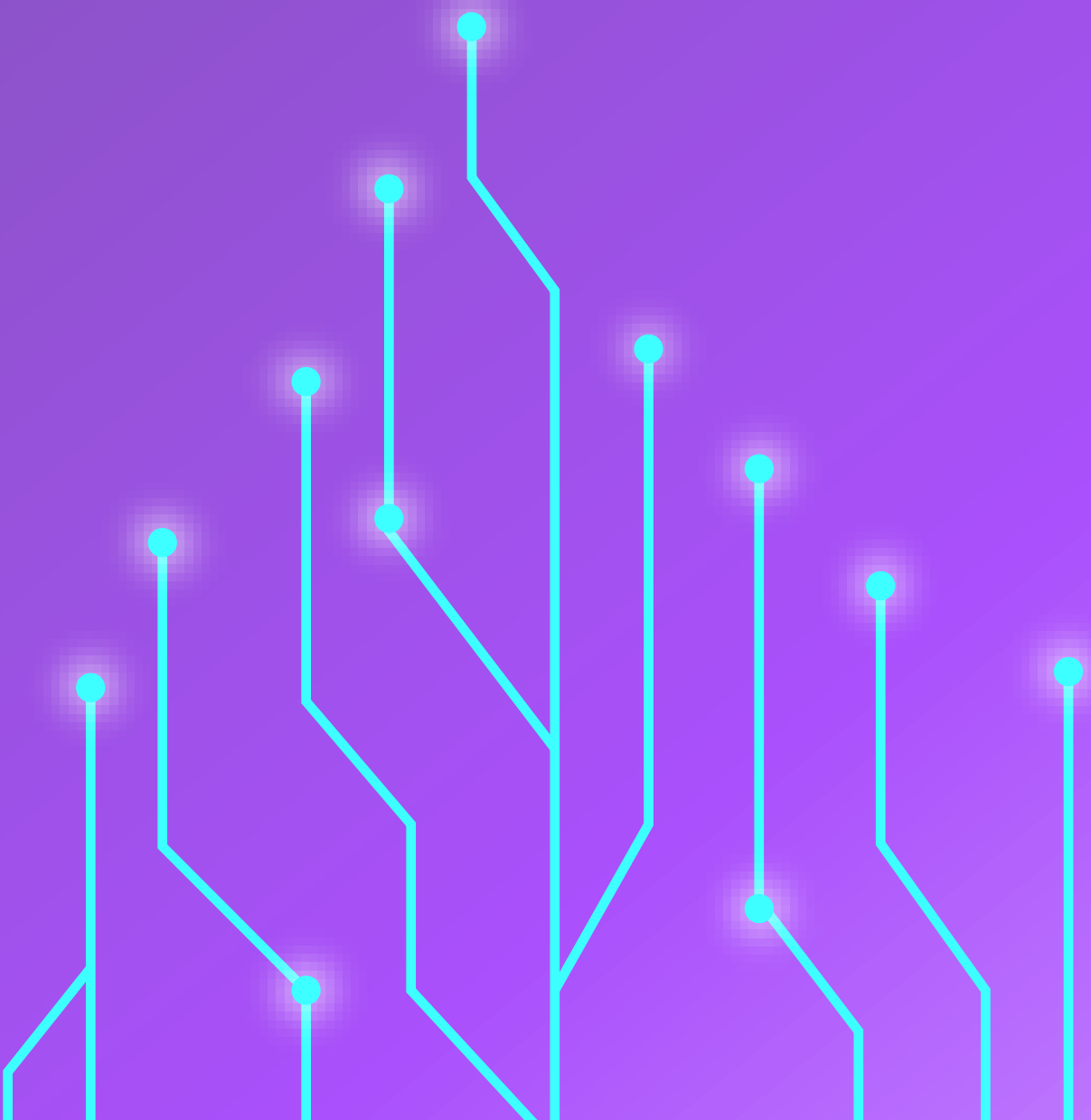
For at løse disse problemer kan eleverne udtænke strategier, der sigter mod at afbøde algoritmisk bias og fremme fairness og retfærdighed i AI-systemer. Det kan omfatte implementering af teknikker til detektering og afbødning af bias i udviklingsfasen, sikring af forskelligartet repræsentation i træningsdata og fremme af gennemsigtighed og ansvarlighed i algoritmiske beslutningsprocesser.

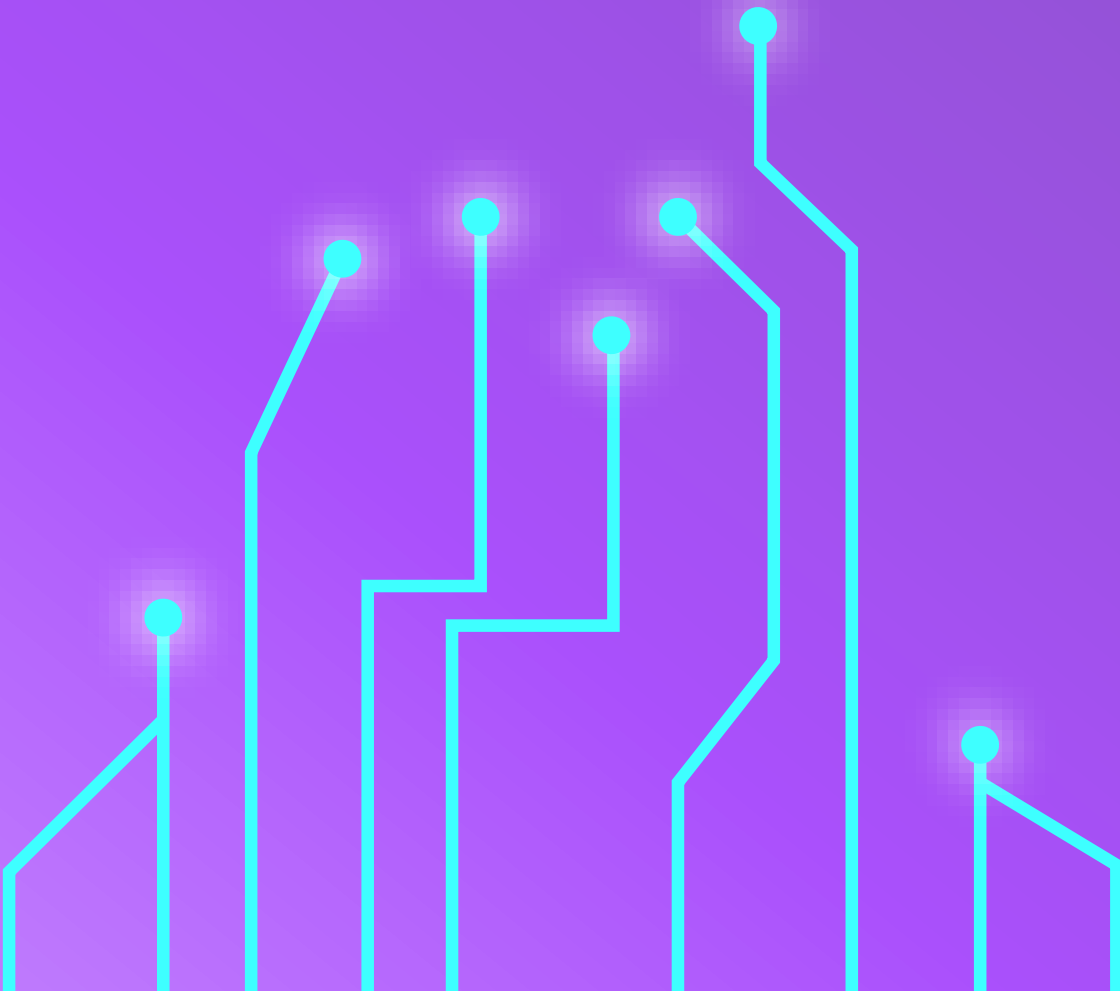
Derudover kan interessenter være fortalere for udvikling og implementering af lovgivningsmæssige rammer og retningslinjer, der prioriterer retfærdighed, ligestillelse og respekt for menneskerettigheder i udviklingen og anvendelsen af kunstig intelligens.



04. Principper for retfærdighed i AI-systemer

CU5 | Menneskerettigheder og retfærdighed





04. Principper for retfærdighed i AI-systemer

Retfærdighed i AI-systemer omfatter en række principper.

Retfærdighed i AI-systemer omfatter en række principper, der er grundlæggende for at sikre retfærdig behandling og resultater for alle individer:

- Et af kerneprincipperne er **lige muligheder**, hvilket betyder, at man skal give folk de samme chancer og muligheder uanset deres baggrund eller karakteristika. Dette princip har til formål at skabe lige vilkår og forhindre diskrimination baseret på faktorer som race, køn eller socioøkonomisk status.
- **Ikke-diskrimination/ligestillelse** er et andet vigtigt princip, der forbyder brugen af vilkårlige eller irrelevante kriterier i beslutningsprocesser. AI-systemer må ikke diskriminere enkeltpersoner eller grupper baseret på beskyttede karakteristika som race, køn, religion eller handicap. I stedet skal de behandle alle individer retfærdigt og upartisk, uanset deres personlige egenskaber.
- **Proceduremæssig** retfærdighed henviser til retfærdigheden i de processer, der bruges til at træffe beslutninger i AI-systemer. Det omfatter gennemsigtighed, ansvarlighed og retten til at appellere eller udfordre beslutninger. Proceduremæssig retfærdighed sikrer, at enkeltpersoner har en stemme i beslutningsprocesser, og at beslutninger træffes på en gennemsigtig og ansvarlig måde.



- Retfærdighed er et princip, der fokuserer på at opnå fairness og retfærdighed ved at tackle systemiske uligheder og give ressourcer og muligheder til dem, der har mest brug for dem. I forbindelse med AI-systemer indebærer retfærdighed, at man designer algoritmer og politikker, der prioriterer marginaliserede eller dårligt stillede gruppers behov og sigter mod at reducere forskellene i adgang til muligheder og ressourcer.
- **Retfærdighed** er et bredere begreb, der omfatter fairness, retfærdighed og beskyttelse af menneskerettigheder. Det søger at sikre, at enkeltpersoner får en retfærdig behandling, og at deres rettigheder og værdighed opretholdes. Retfærdighed i AI-systemer kræver overholdelse af etiske standarder, lovbestemmelser og samfundsnormer, der fremmer lighed, retfærdighed og respekt for menneskerettighederne.

Det er vigtigt at skelne mellem disse principper for at forstå de nuancerede etiske overvejelser, der er involveret i udvikling og anvendelse af kunstig intelligens. Mens lige muligheder fokuserer på at sikre lige chancer for alle individer, ikke-diskrimination/ligestillelse forbyder uretfærdig behandling baseret på personlige karakteristika. Proceduremæssig retfærdighed lægger vægt på gennemsigtighed og ansvarlighed i beslutningsprocesser, mens retfærdighed sigter mod at tackle systemiske uligheder og fremme retfærdighed for marginaliserede grupper. Endelig omfatter retfærdighed bredere etiske og juridiske overvejelser i forbindelse med menneskerettigheder og samfundets velfærd.



Anvendelse af retfærdighedsprincipper i AI-udvikling kræver en proaktiv og flerdimensionel tilgang, der tager højde for de etiske, sociale og juridiske konsekvenser af AI-teknologier. Disse principper kan integreres i AI-systemer for at fremme retfærdige resultater og mindske algoritmisk bias og dermed lægge grunden til en mere retfærdig og rimelig fremtid for de kommende generationer.

Et vigtigt aspekt ved at anvende retfærdighedsprincipper i AI-udvikling er at sikre, at algoritmer designs og trænes ved hjælp af forskellige og repræsentative datasæt. Ved at inddrage forskellige perspektiver og erfaringer i udviklingsprocessen kan udviklere mindske risikoen for forudindtagede resultater og sikre, at AI-systemer er retfærdige og inkluderende.

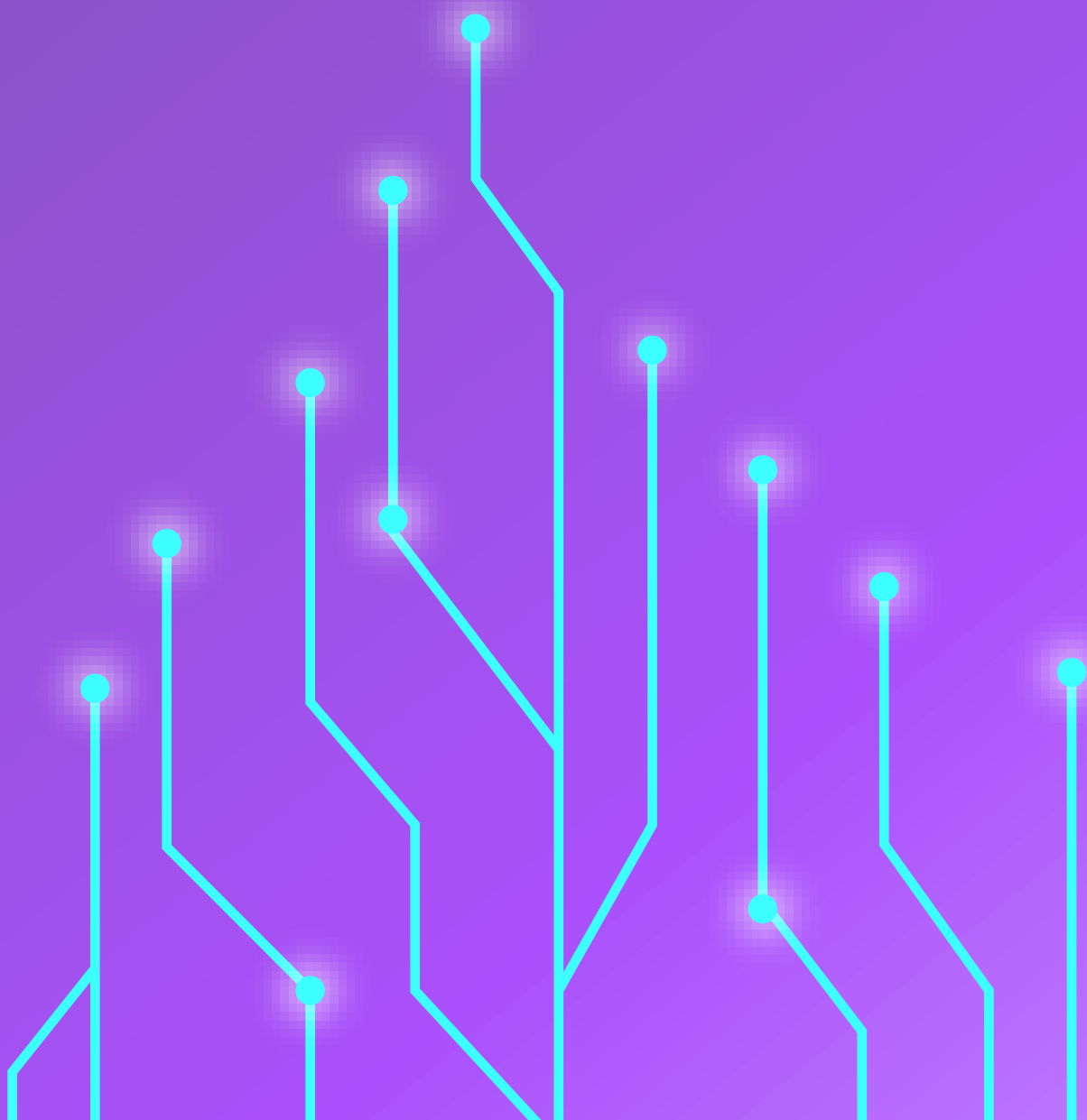
Gennemsigtige AI-systemer gør det muligt for interessenter at forstå, hvordan beslutninger træffes, og at holde udviklere ansvarlige for deres handlinger. Desuden sikrer mekanismer for klageadgang og oprejsning, at personer, der er berørt af forudindtagede AI-systemer, har mulighed for at søge retfærdighed og afhjælpning.

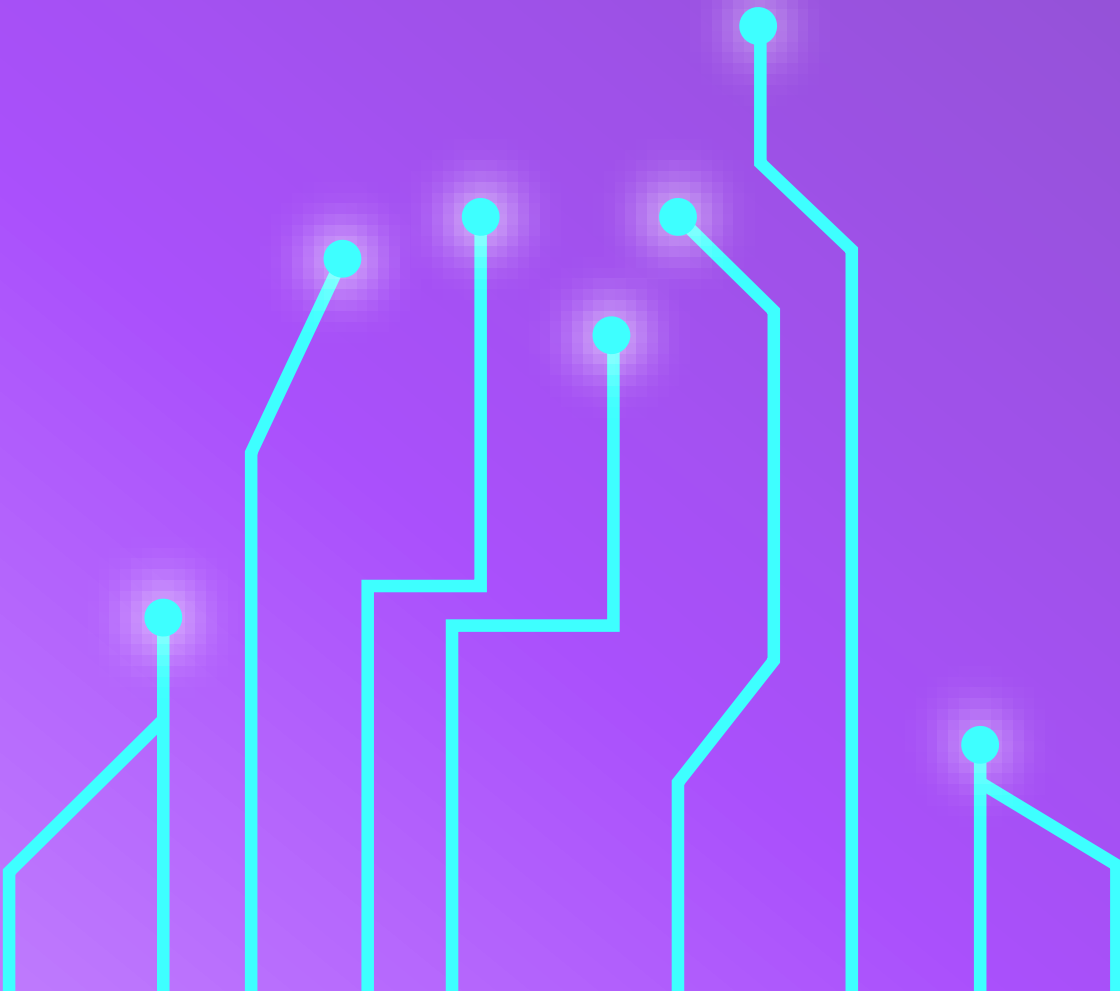
Desuden er strategier til fremme af retfærdighed i AI-systemer, såsom design af algoritmer, der prioriterer retfærdighed og inklusivitet og sigter mod at reducere forskelle i adgang til muligheder og ressourcer. Ved at overveje marginaliserede eller dårligt stillede gruppers behov og perspektiver kan udviklere sikre, at AI-teknologier tjener alle individers interesser og bidrager til et mere retfærdigt samfund.



05. Konklusion

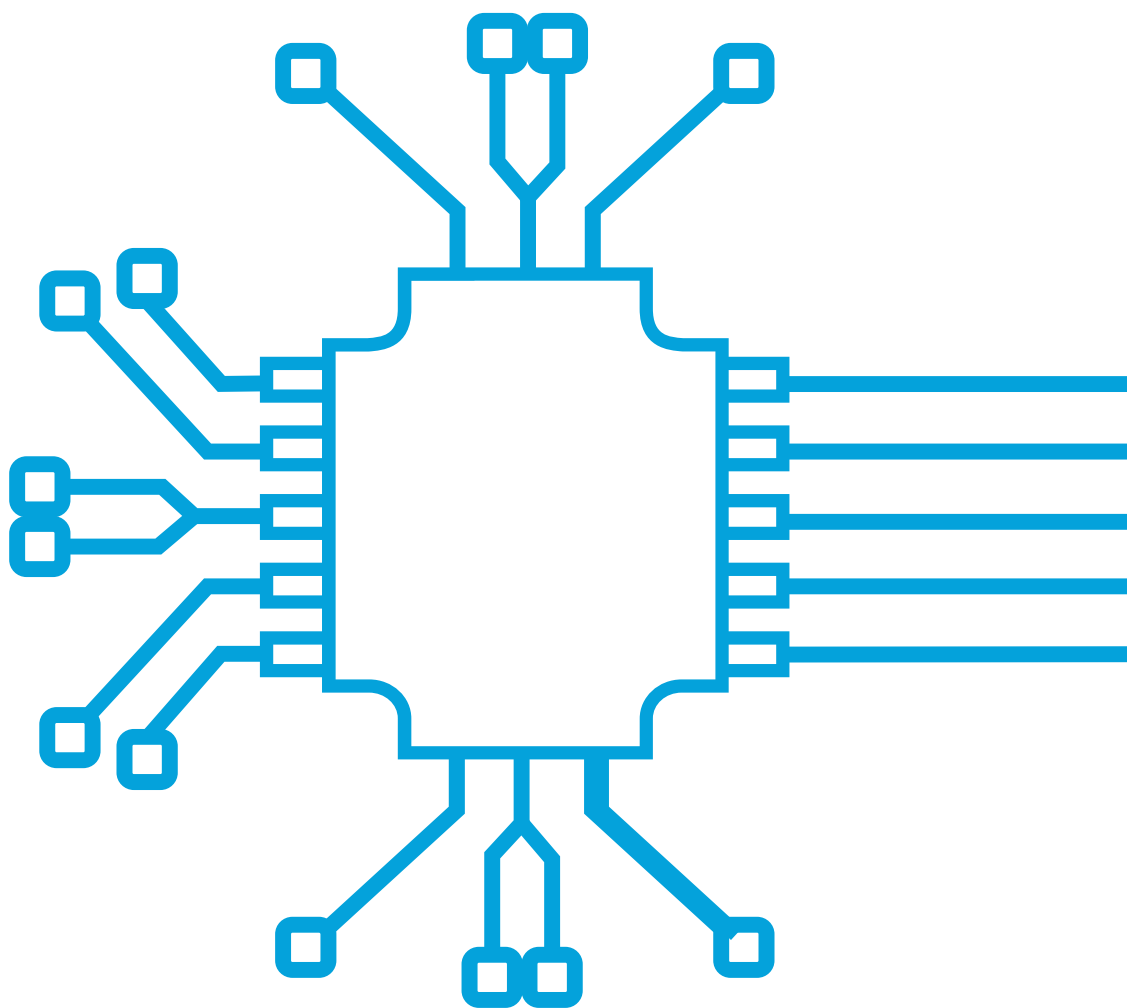
CU5 | Menneskerettigheder og retfærdighed

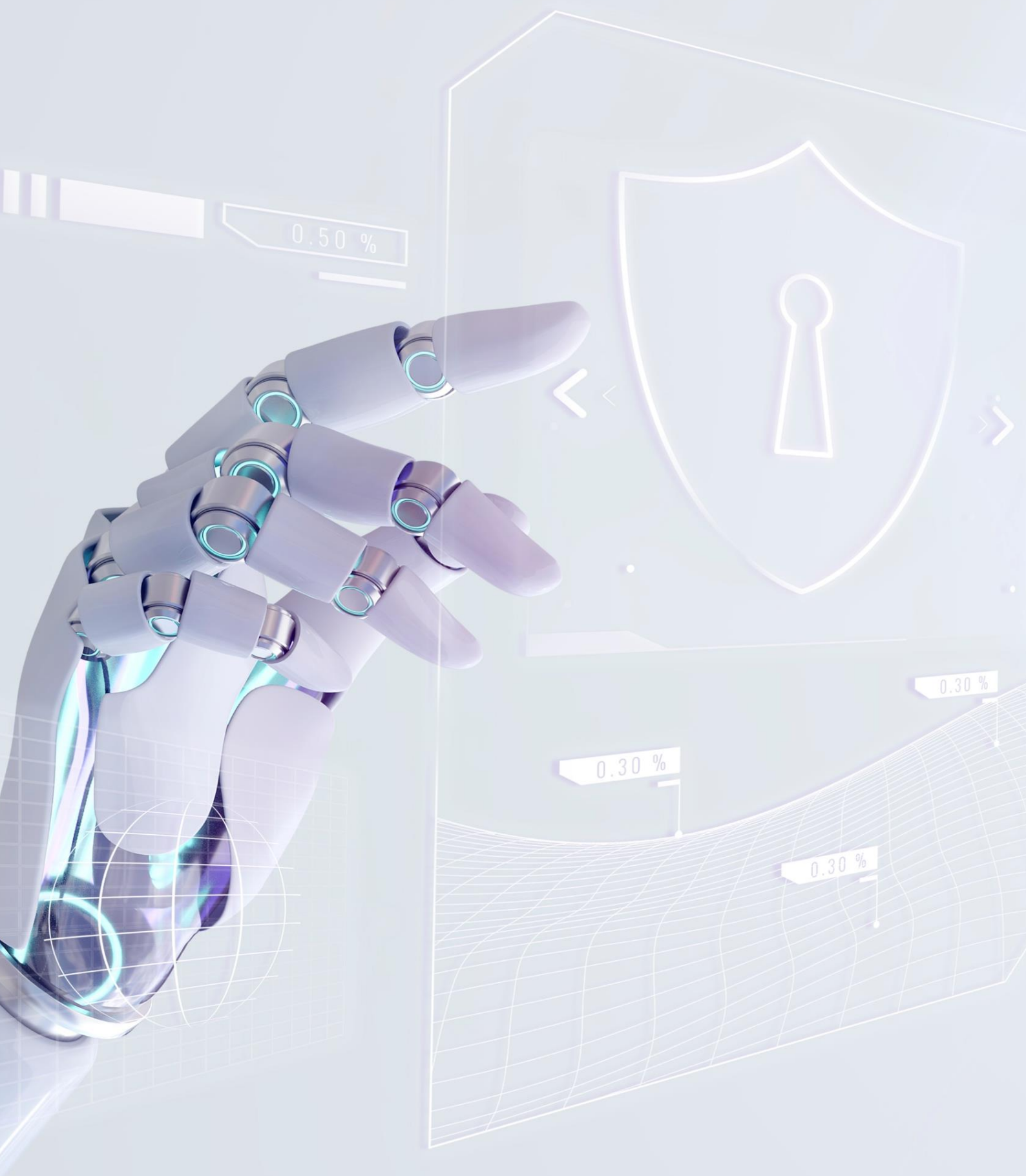




05. Konklusion

Konklusionen er, at forståelse og anvendelse af principper for menneskerettigheder og retfærdighed i AI-systemer er afgørende for at fremme retfærdige resultater og beskytte grundlæggende rettigheder i den digitale tidsalder.







Charlæ



Universitat
de les Illes Balears



ENGAGING PEOPLE



INNOVATION TRAINING CENTER



AARHUS UNIVERSITY



VAAKAN AMMATTIKORKEAKOULU
UNIVERSITY OF APPLIED SCIENCES



Medfinansieret af
Den Europæiske

Finansieret af Den Europæiske Union. Synspunkter og holdninger, der kommer til udtryk, er udelukkende forfatterens/forfatternes og er ikke nødvendigvis udtryk for Den Europæiske Unions eller Det Europæiske Forvaltningsorgan for Uddannelse og Kulturs (EACEA) officielle holdning. Hverken den Europæiske Union eller



2022-1-ES01-KA220-HED-000085257