



# Etisk AI-mikrocertifikat

HÆFTE

CU3 | Ansvarlighed

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.

**ANSVARSFRAKRIVELSE:** 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. Definition af ansvarlighed og opdyrkning af ansvar i AI**

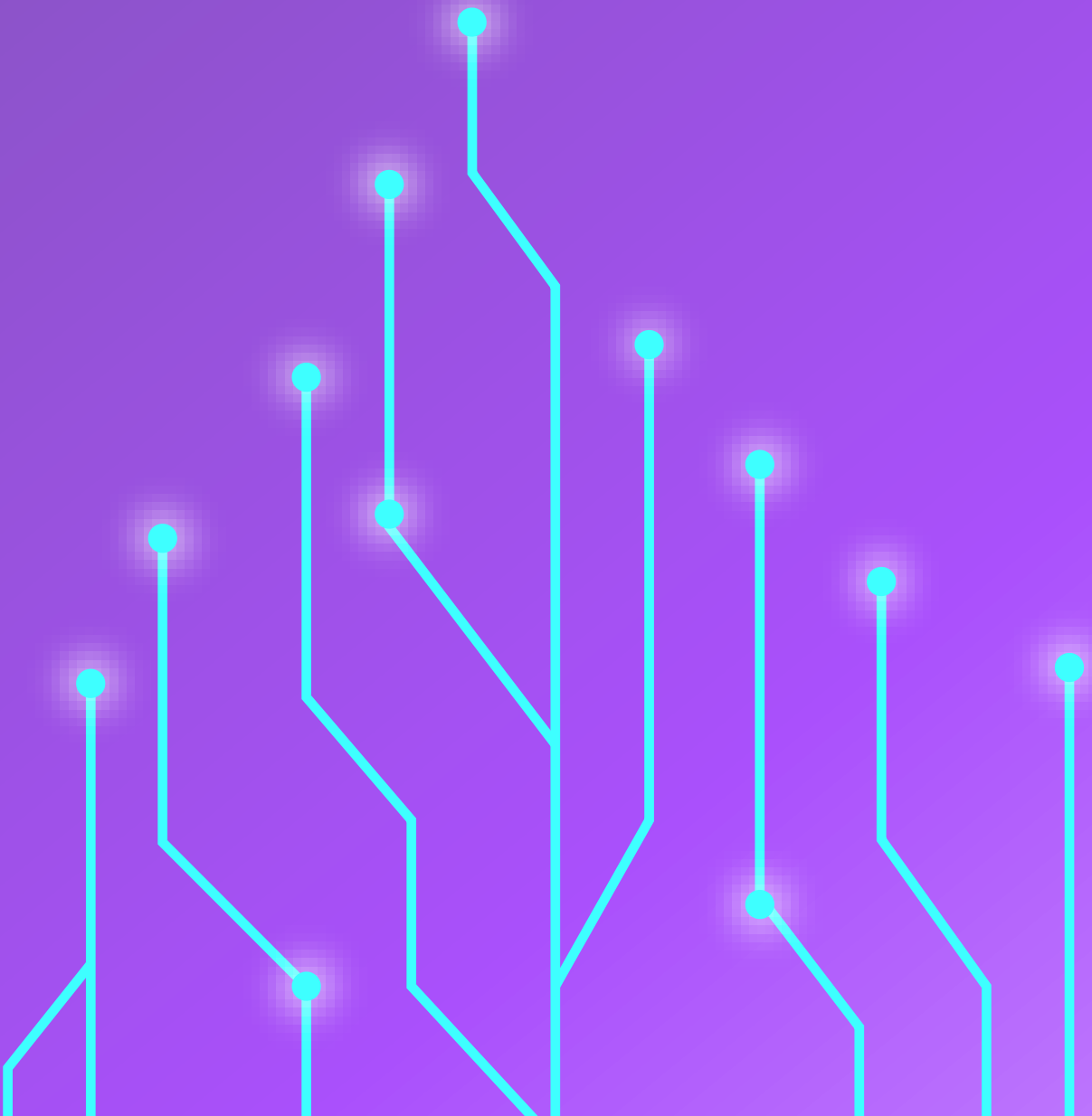
**03. Ansvarligheds rolle i håndteringen af Algoritmisk bias**

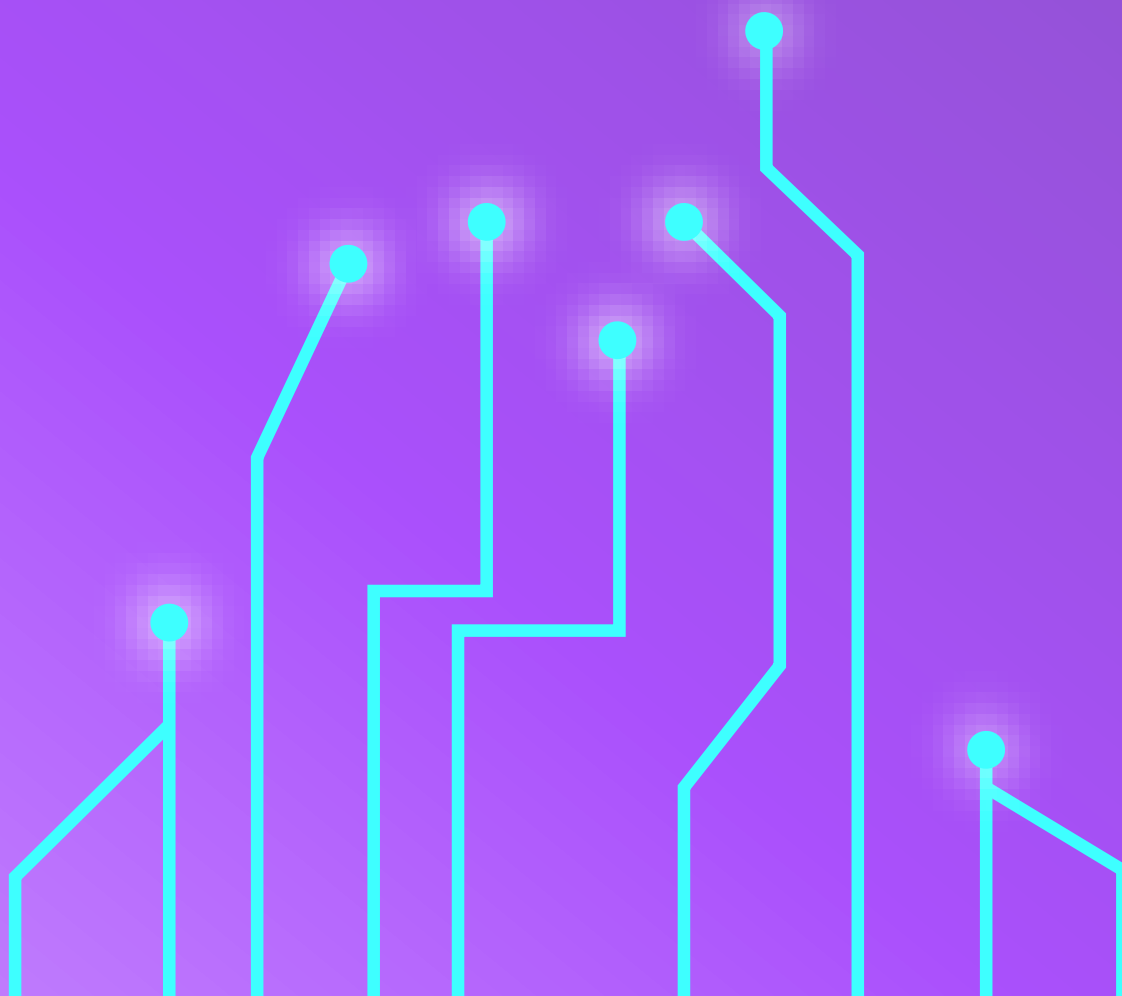
**04. Mekanismer til sikring af ansvarlighed i AI-systemer**

**05. Konklusion**

# 01. Introduktion

CU3 | Ansvarlighed





## 01. Introduktion

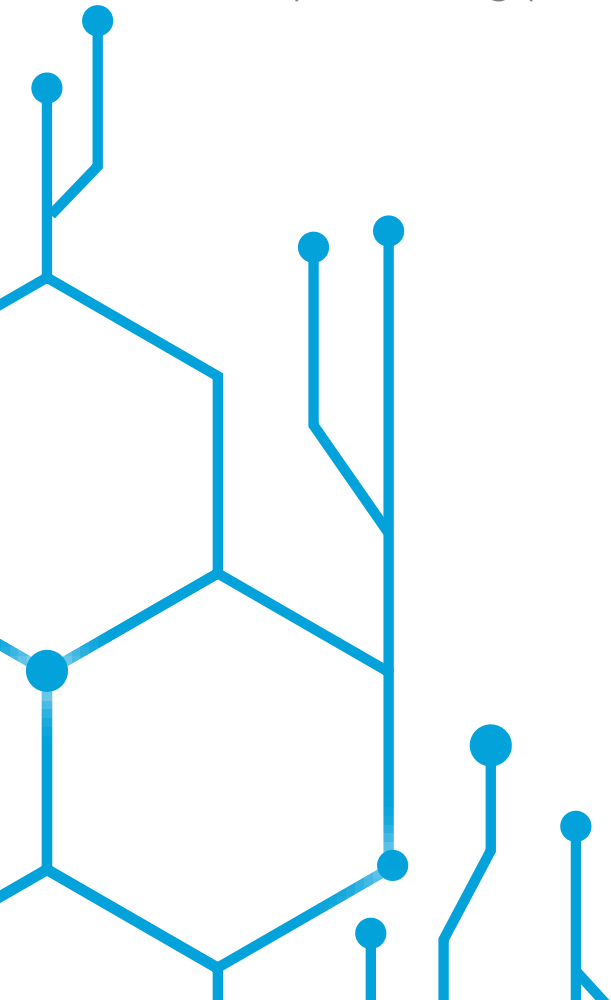
I denne kompetenceenhed vil eleverne tilegne sig grundlæggende viden om ansvarlighed i AI med fokus på at forstå grundlæggende koncepter, identificere AI-udvikleres og -brugeres ansvar for at sikre etiske AI-systemer med minimal skade og anerkende de virkelige konsekvenser af at vedtage og implementere mekanismer, der fremmer ansvarlighed i AI-systemer.

Vidensmålene for denne enhed omfatter:

- **Definition af ansvarlighed og dyrkelse af ansvarlighed i AI:** Denne enhed fokuserer på det grundlæggende begreb ansvarlighed i AI, som indebærer en forventning om, at designere, udviklere og brugere overholder standarder og lovgivning for at sikre, at AI fungerer korrekt i hele deres livscyklus (Fjeld et al., 2020). AI-udvikleres og -brugeres ansvar for at fremme etiske AI-systemer med minimal skade vil blive præsenteret, og det vil blive fremhævet, hvor vigtigt det er at opdyrke en ansvarsfølelse for etisk udvikling og brug af AI og forstå betydningen af ansvarlighed for at minimere skade og fremme etisk adfærd i AI-systemer.

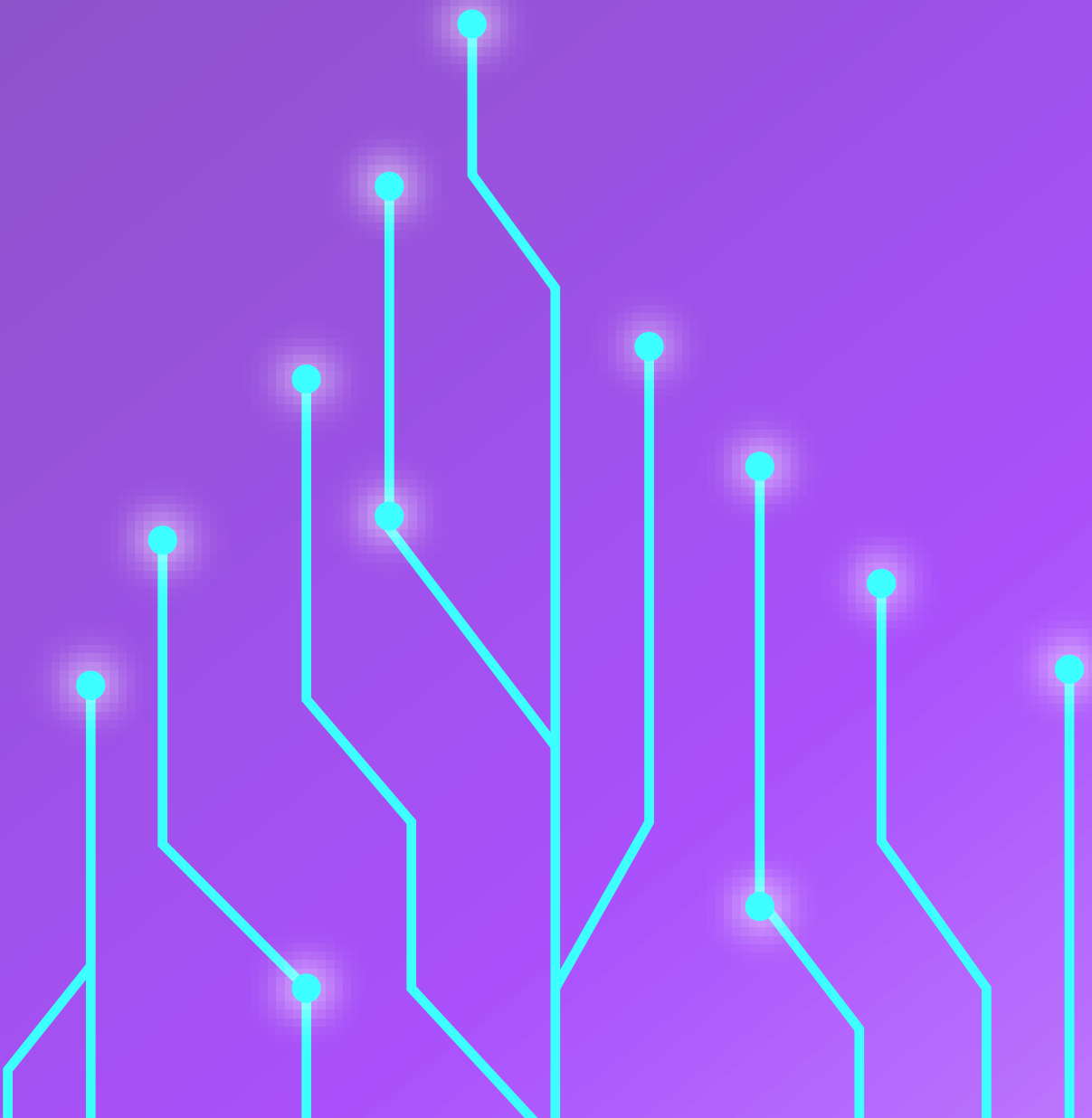


- **Accountability's Role in Addressing Algorithmic Bias**, der anerkender den centrale rolle, som ansvarlighed spiller i at forebygge og afbøde virkningerne af algoritmisk bias. Vi vil udforske forholdet mellem ansvarlighed og algoritmisk bias og værdsætte vigtigheden af at påtage sig ansvaret for AI-systemer for at forbedre beslutningstagningen og fremme mere retfærdige resultater.
- **Implementering af mekanismer for ansvarlighed i AI-systemer** for at blive bekendt med forskellige mekanismer, der har til formål at sikre ansvarlighed i AI-systemer. Vi vil udforske vigtigheden af effektiv kommunikation med interessenter, som f.eks. at give information om rationalet og logikken i AI-systemer, herunder inputs, outputs og beslutningsprocesser. Eksempler på sådanne mekanismer omfatter retningslinjer, netikette, politikker for acceptabel brug (AUP) og regler for AI-udviklere og -brugere.



# 02. Definition af ansvarlighed og opdyrkning af ansvar i AI

CU3 | Ansvarlighed







## 02. Definér ansvarlighed og dyrk den ansvar i AI

Ansvarlighed er afgørende for en etisk udvikling og anvendelse af systemer med kunstig intelligens (AI).

Det fungerer som et værn mod uetisk praksis og sikrer, at AI-teknologier fungerer i overensstemmelse med etablerede standarder, regler og etiske rammer. Dette afsnit dykker ned i begrebet ansvarlighed i AI og undersøger dets principper, mekanismer og konsekvenser i nutidens teknologiske landskaber.

Kernen i ansvarlighed inden for AI er, at designere, udviklere og brugere er ansvarlige for at opretholde etiske standarder i hele AI-systemets livscyklus. Det indebærer gennemsigtighed i beslutningstagningen, sporbarhed af algoritmiske output og lydhørhed over for feedback fra interessenter. Desuden kræver ansvarlighed proaktiv risikoidentifikation og -reduktion for at fremme etisk udvikling og brug af AI.

Juridiske rammer, industristandarder og etiske retningslinjer former landskabet for AI-ansvarlighed. Juridisk ansvarlighed krydses med lovgivningsmæssige krav og ansvarsovervejelser, der pålægger forpligtelser til at overholde databeskyttelseslove og forbrugerbeskyttelsesregler.



Branchestandarder og etiske retningslinjer giver rammer for at navigere i etiske dilemmaer og fremme ansvarlig AI-praksis.

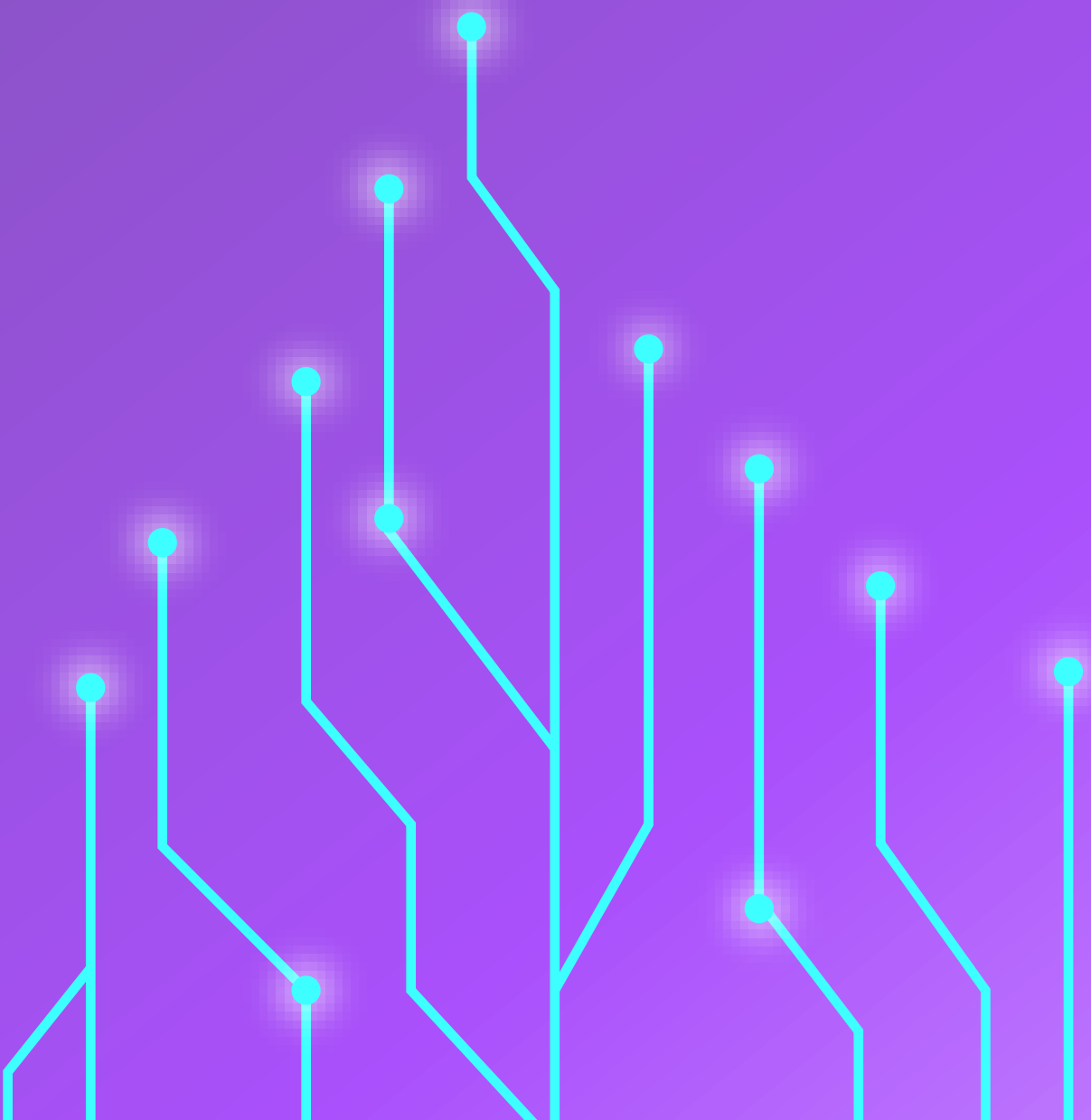
Algoritmisk ansvarlighed indebærer en vurdering af AI-algoritmernes tekniske nøjagtighed og en forståelse af deres socioøkonomiske konsekvenser og etiske implikationer. Bias i AI-algoritmer kan føre til forskellige resultater, fastholde uligheder og underminere tilliden. Derfor kræver ansvarlighed årvågenhed i forhold til at opdage og afbøde bias sammen med mekanismer til gennemsigtighed og forklaring.

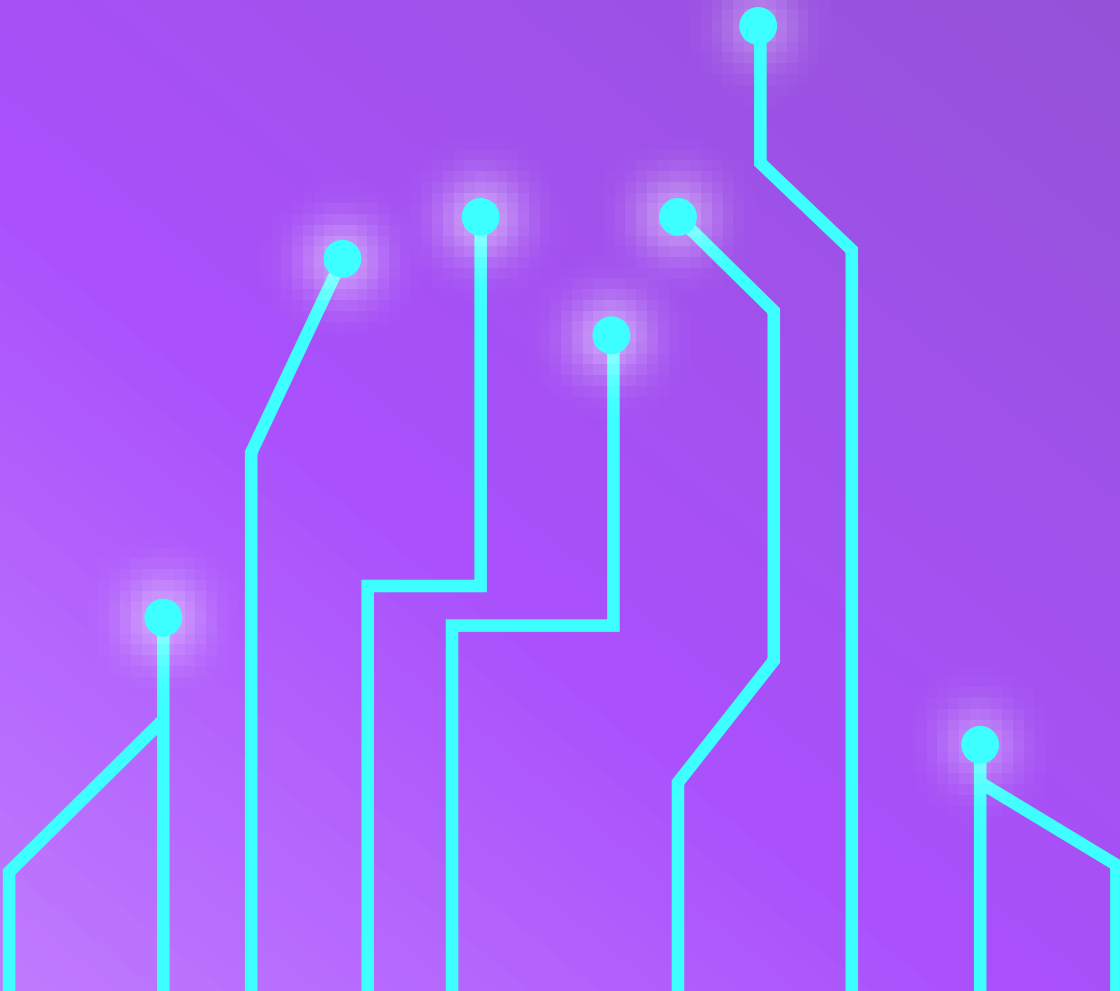
Gennemsigtighed og forklarlighed øger tilliden til og ansvarligheden i AI-systemer og gør det muligt for interessenter at forstå algoritmiske beslutningsprocesser og resultater. Etisk ansvarlighed kræver, at AI-udvikling tilpasses etiske principper, menneskerettigheder og samfundsmæssige værdier. Samfundsmæssig ansvarlighed indebærer at engagere interessenter for at sikre, at AI-teknologier respekterer menneskelig værdighed, autonomi og retfærdighed.



# 03. Ansvarligheds rolle i håndteringen af Algoritmisk bias

CU3 | Ansvarlighed





### 03. Ansvarligheds rolle i håndteringen af Algoritmisk bias

Ansvarlighed spiller en central rolle i håndteringen af algoritmisk bias i systemer med kunstig intelligens (AI).

Efterhånden som AI-teknologier bliver mere og mere integreret i forskellige aspekter af samfundet, er bekymringer om bias og retfærdighed kommet i forgrunden. Algoritmisk bias refererer til systematiske fejl eller uretfærdighed i AI-algoritmer, der resulterer i diskriminerende resultater for visse personer eller grupper. Disse skævheder kan fastholde eksisterende uligheder, forstærke stereotyper og underminere tilliden til AI-systemer. I dette afsnit vil vi undersøge, hvordan ansvarlighed kan hjælpe med at identificere, afbøde og forhindre algoritmisk bias og sikre, at AI-teknologier er fair, etiske og retfærdige.

Et af de vigtigste aspekter af ansvarlighed i forbindelse med algoritmisk bias er **gennemsigtighed**. Gennemsigtighed henviser til AI-systemers åbenhed og tilgængelighed, så interessenter kan forstå, hvordan algoritmer fungerer, hvorfor visse beslutninger træffes, og hvilke faktorer der påvirker deres output. Gennemsigtige AI-systemer giver brugerne mulighed for at granske og udfordre algoritmiske resultater, hvilket fremmer tillid og ansvarlighed.



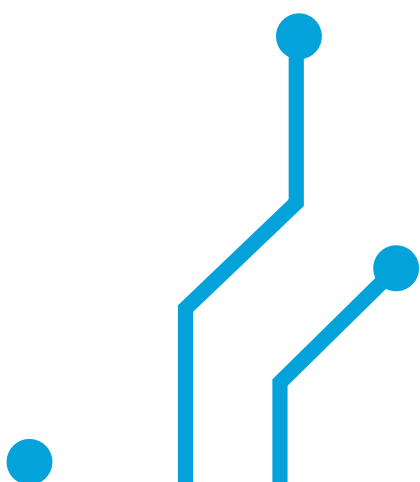


Hvis et AI-system f.eks. bruges til automatiseret beslutningstagning i forbindelse med lånegodkendelser, vil gennemsigtighed indebære offentliggørelse af de kriterier og faktorer, der er taget i betragtning i beslutningsprocessen, såsom kredithistorik, indkomst og demografiske oplysninger. Denne gennemsigtighed gør det muligt for enkeltpersoner at forstå, hvorfor deres låneansøgning blev godkendt eller afvist, og giver mulighed for at klage, hvis de mener, at beslutningen var forudindtaget eller uretfærdig.

Et andet aspekt af ansvarlighed i håndteringen af algoritmisk bias er **forklarbarhed**. Forklarlighed henviser til AI-systemers evne til at give fortolkelige forklaringer på deres beslutninger og handlinger.

Forklarende AI gør det muligt for brugerne at forstå rationalet bag algoritmiske resultater og at identificere og korrigere bias eller fejl.

Hvis et AI-drevet ansættelsesværktøj f.eks. automatisk afviser jobansøgninger fra personer med bestemte demografiske karakteristika, vil forklarbarheden indebære, at der gives forklaringer på, hvorfor disse ansøgninger blev afvist, f.eks. ved at fremhæve specifikke kriterier eller funktioner, der førte til beslutningen. Denne forklarlighed giver brugerne mulighed for at udfordre forudindtagede beslutninger og træffe korrigerende foranstaltninger for at mindske bias i AI-systemer.



Ansvarlighed indebærer også **ansvar for design, udvikling og implementering af AI-systemer**. AI-udviklere og -udøvere har et ansvar for at sikre, at deres systemer er fair, upartiske og retfærdige. Dette ansvar omfatter grundig testning og validering for at identificere og afbøde bias i AI-algoritmer samt overvågning og revision af AI-systemer for at sikre løbende overholdelse af etiske og juridiske standarder. AI-udviklere kan f.eks. bruge teknikker som bias-detektionsalgoritmer, fairness-metrikker og kontradiktorisk testning til at identificere og håndtere bias i deres modeller. De kan også implementere processer til regelmæssig revision og gennemgang af AI-systemer for at identificere og håndtere bias, der kan opstå over tid.

Desuden kræver ansvarlighed i håndteringen af algoritmisk bias **samarbejde og engagement med forskellige interessenter**. Det omfatter inddragelse af berørte samfund, interessegrupper, politiske beslutningstagere og tilsynsorganer i design, udvikling og tilsyn med AI-systemer. Ved at samarbejde med interessenter kan AI-udøvere få værdifuld indsigt i de potentielle konsekvenser af deres teknologier og sikre, at deres systemer designes og implementeres på en måde, der respekterer alle individers og gruppers rettigheder og interesser. Hvis et AI-system f.eks. bruges til forudseende politiarbejde, vil ansvarlighed indebære at konsultere medlemmer af lokalsamfundet og borgerrettighedsorganisationer for at forstå deres bekymringer og perspektiver på brugen af AI i retshåndhævelsen og for at håndtere eventuelle fordomme eller skader.

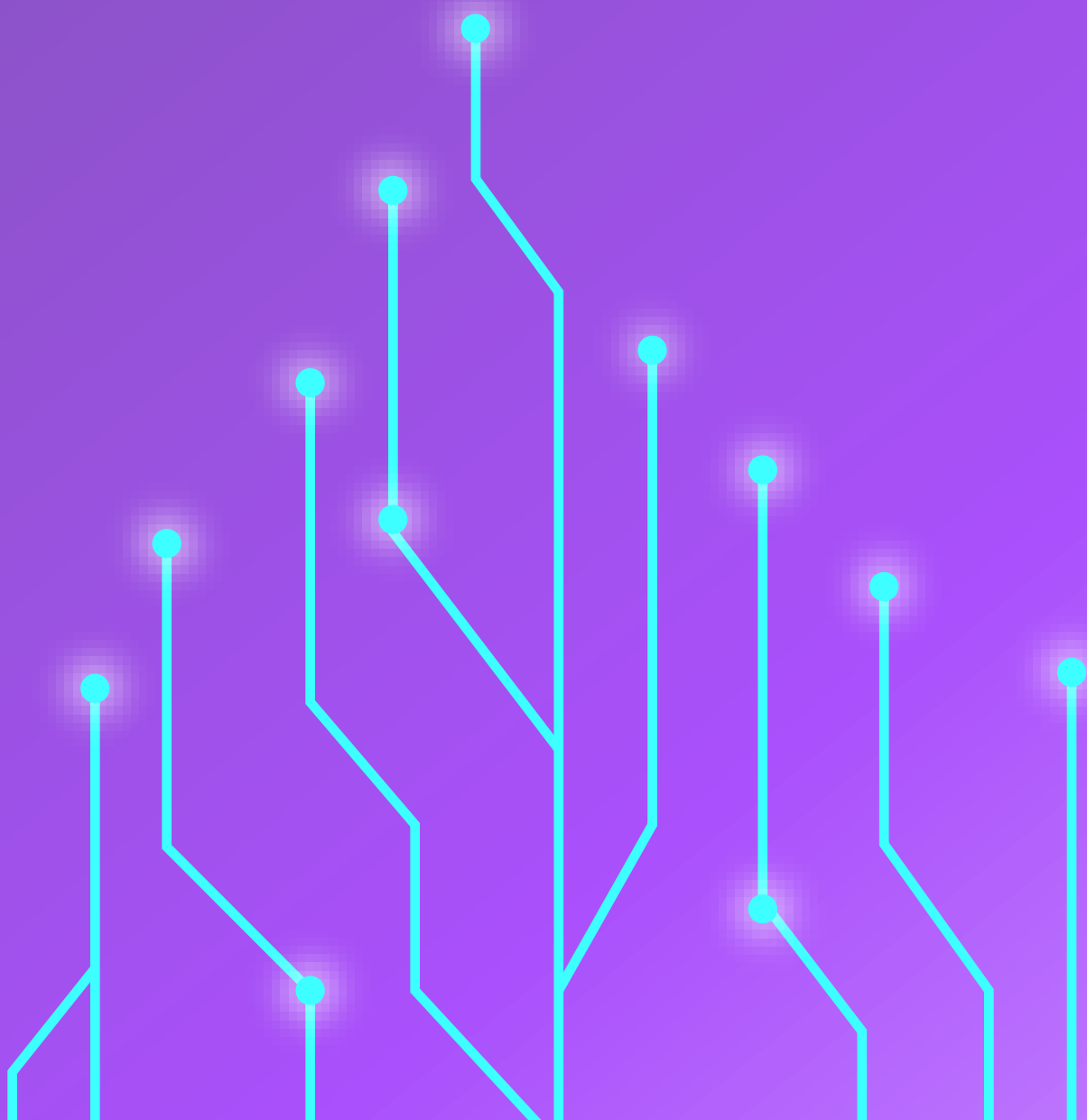


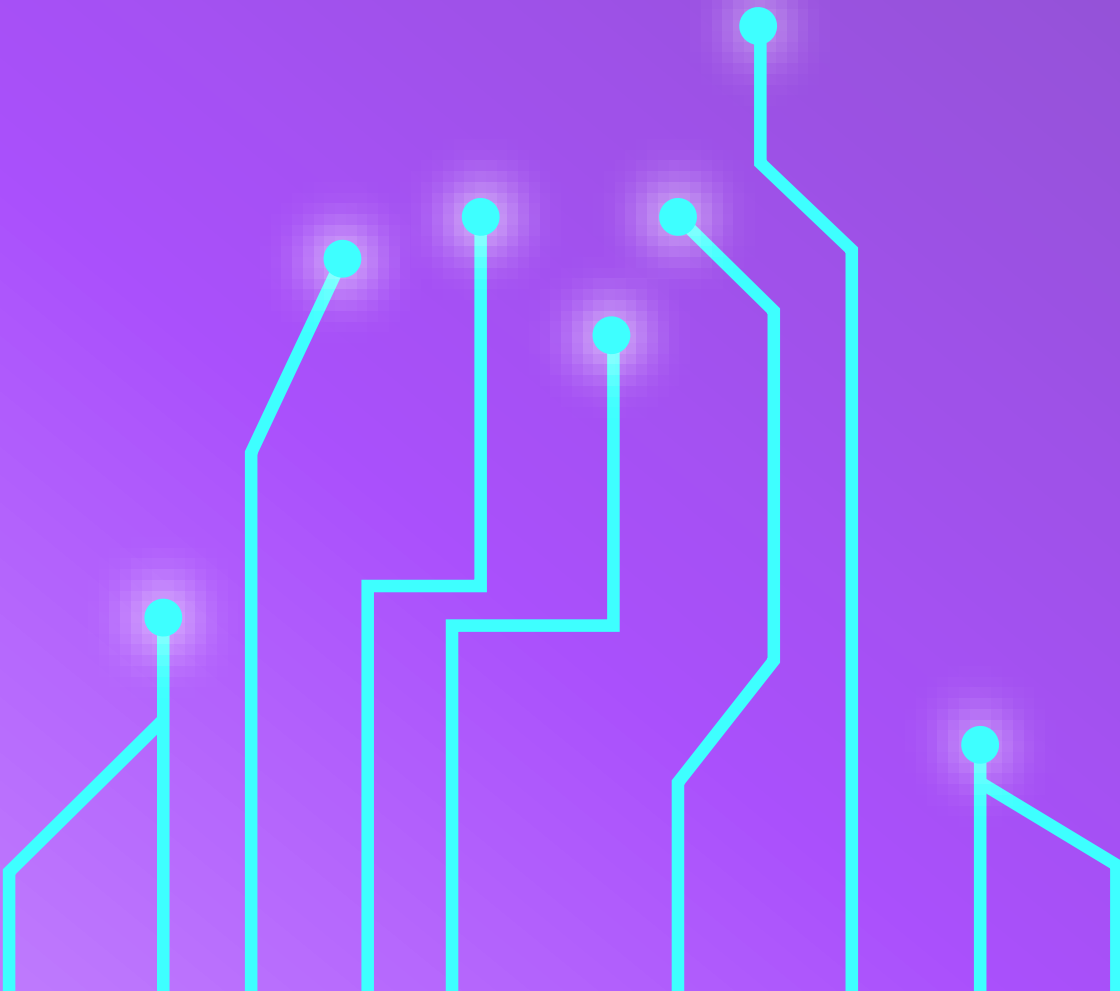


# AI

# 04. Mekanismer til sikring af ansvarlighed i AI-systemer

CU3 | Ansvarlighed





## 04. Mekanismer til sikring af ansvarlighed i AI-systemer

I dagens hurtigt udviklende teknologiske landskab, hvor systemer med kunstig intelligens (AI) har stor indflydelse på beslutningsprocesser, er det altafgørende at sikre ansvarlighed. Mekanismer til at sikre ansvarlighed i AI-systemer omfatter en række værktøjer og rammer, der er designet til at fremme etisk adfærd, gennemsigtighed og retfærdighed i udviklingen og brugen af AI-teknologier:

- **Retningslinjerne** fungerer som overordnede principper og bedste praksis for AI-udviklere og -brugere og giver vejledning i etiske overvejelser, ansvarlig datahåndtering og strategier for risikominimering.
- Netikette-principper kan tilpasses til at fremme etisk adfærd i AI-systemer, hvor der lægges vægt på respekt, høflighed og ansvarlig kommunikation mellem brugerne.
- **Politikker for acceptabel brug (AUP'er)** skitserer adfærdsstandards og retningslinjer for brug af AI-systemer og beskriver acceptabel adfærd og konsekvenser af forkert brug eller misbrug.

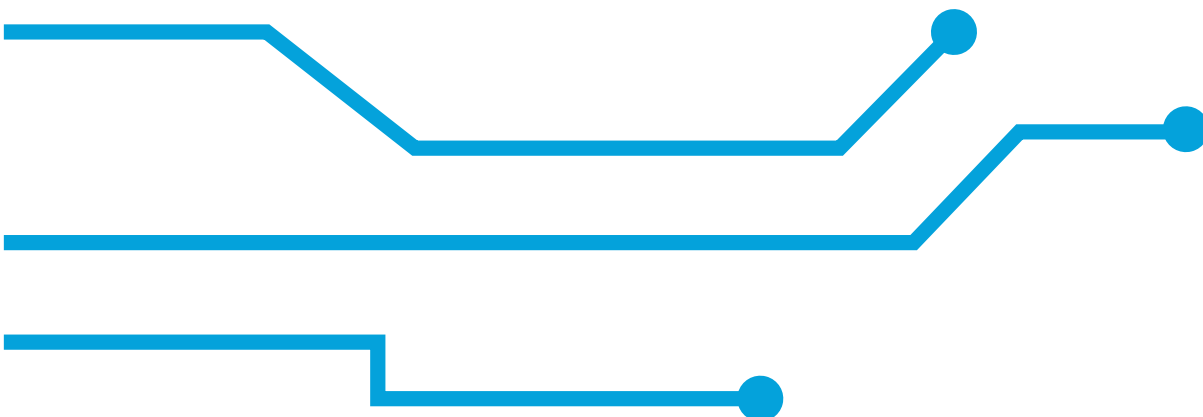




- **Lovgivningsmæssige rammer** fastlægger juridiske krav og standarder for udvikling og anvendelse af AI, herunder bestemmelser om ansvarlighed, gennemsigtighed og tilsyn. I Den Europæiske Union (EU) fastlægger **lovgivningsmæssige rammer som** den **generelle databeskyttelsesforordning (GDPR)** og den foreslåede **lov om kunstig intelligens** juridiske krav og standarder for udvikling og anvendelse af **kunstig intelligens**. Disse regler omfatter bestemmelser om ansvarlighed, f.eks. krav om gennemsigtighed, retfærdighed og tilsyn med udviklingen og brugen af AI-systemer. I henhold til GDPR skal udviklere f.eks. sikre gennemsigtighed i, hvordan AI-systemer behandler persondata, og give enkeltpersoner oplysninger om den logik, der er involveret i automatiseret beslutningstagning.

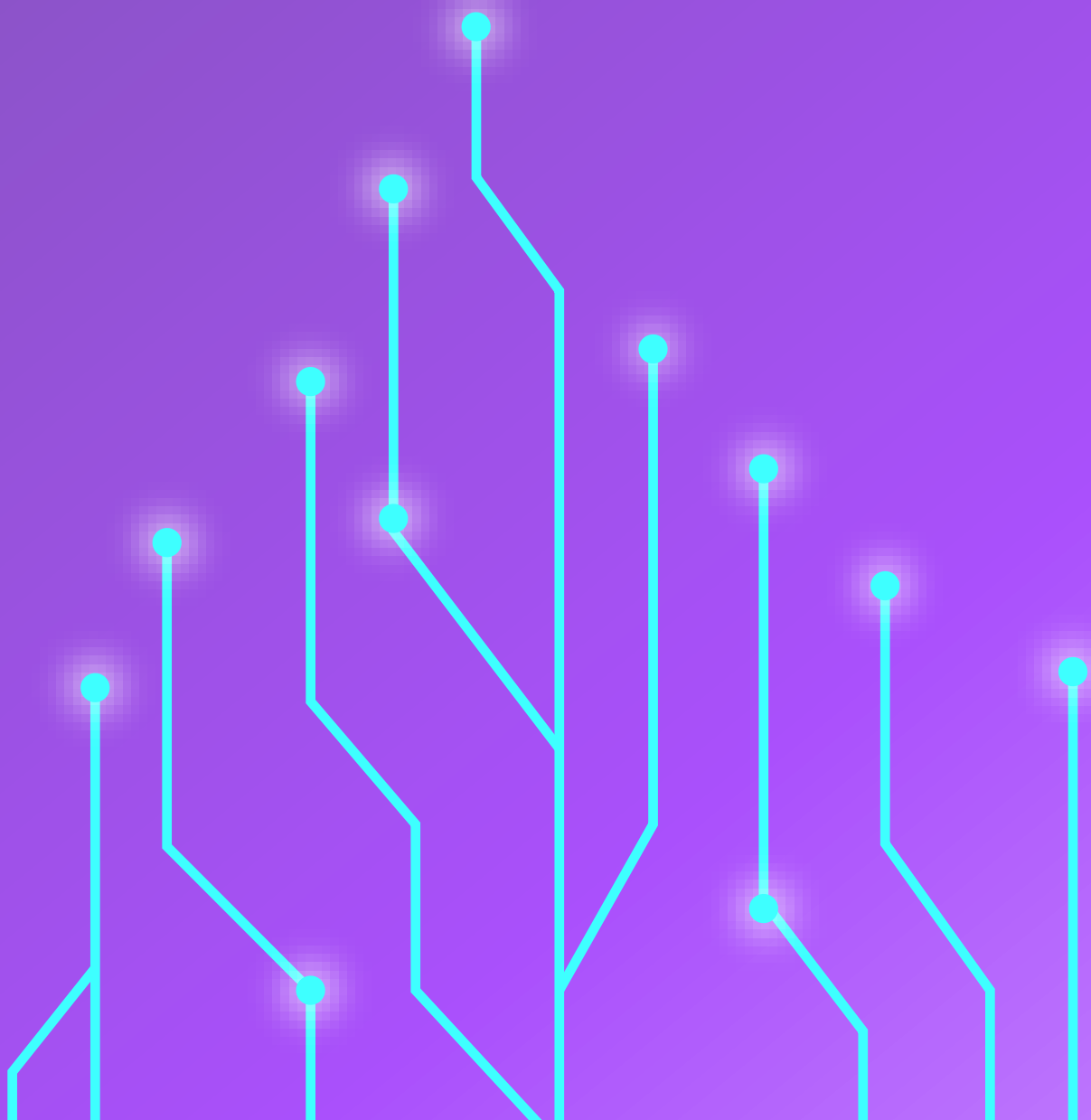


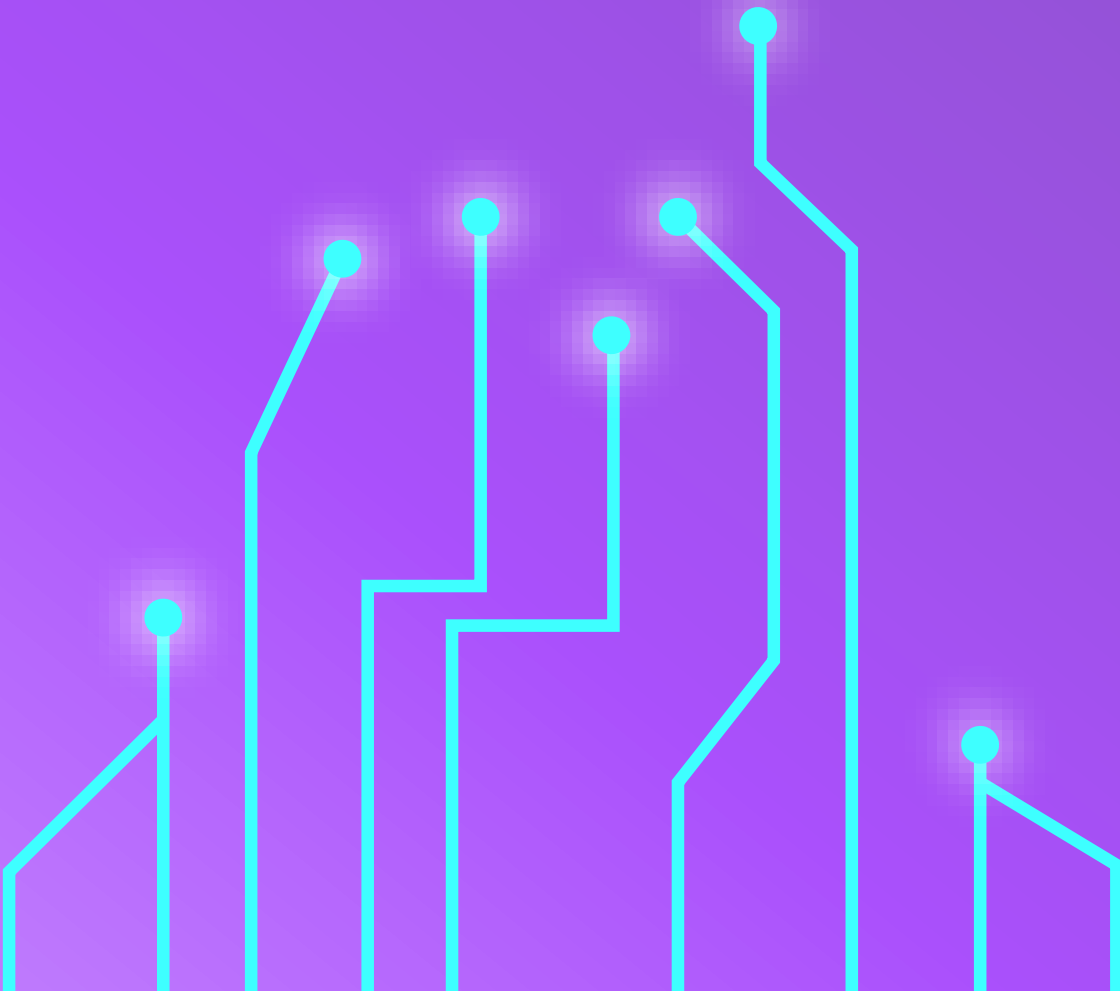
Ved at udnytte disse mekanismer kan interessenter fremme en kultur med ansvarlighed og etisk adfærd inden for udvikling og anvendelse af AI og sikre, at AI-teknologier bruges ansvarligt og etisk til gavn for samfundet.



# 05. Konklusion

CU3 | Ansvarlighed





## 05. Konklusion

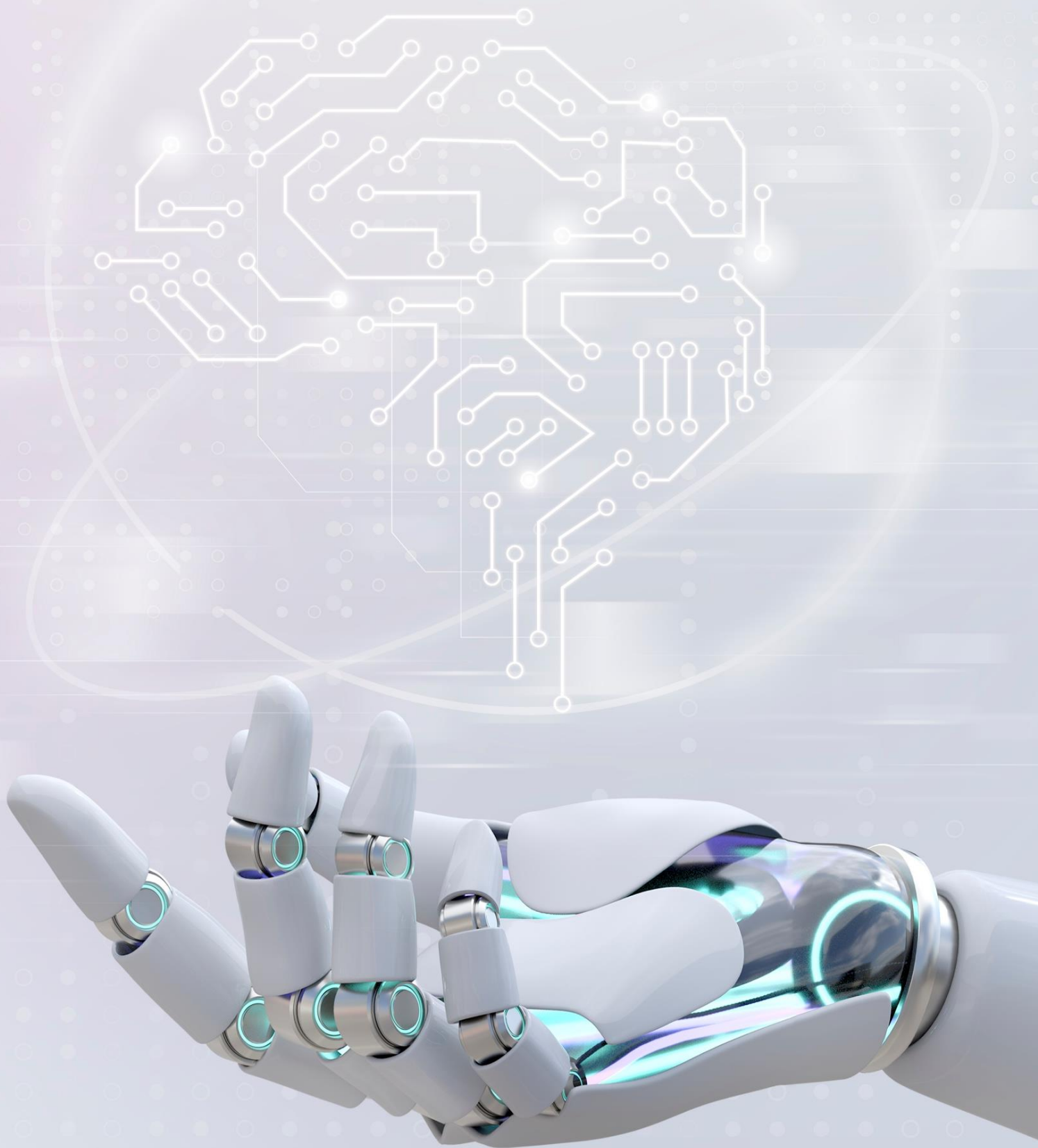
Konklusionen er, at ansvarlighed er en hjørnesteen i den etiske udvikling og implementering af systemer med kunstig intelligens (AI). I denne CU har vi udforsket den mangesidede karakter af ansvarlighed i AI ved at dykke ned i dens grundlæggende principper, underliggende mekanismer og praktiske implikationer. Fra gennemsigtighed og forklarlighed til inddragelse af interessenter og overholdelse af lovgivningen fungerer ansvarlighed som et vejledende princip for at sikre, at AI-teknologier udvikles og anvendes på en måde, der opretholder etiske standarder, fremmer gennemsigtighed og fremmer retfærdighed og lighed.

Ved at tage ansvar kan interessenter navigere i AI-teknologiernes kompleksitet med integritet, ansvar og en forpligtelse til samfundets velbefindende.

Efterhånden som AI fortsætter med at udvikle sig og gennemtrænge forskellige aspekter af menneskelivet, er ansvarlighed fortsat afgørende for at skabe en fremtid, hvor AI-teknologier tjener det fælles bedste og opretholder værdierne etik, retfærdighed og menneskelig værdighed.









# 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