

Kursus i algoritmiske grundprincipper og etik i AI: fra teori til praksis

Værktøjer til synkrone sessioner

CU5 | Casestudier og projekter
Understøttende PowerPoint-slides

INDHOLD

- INDLEDNING - 3
- ALGORITMISK OG MASKINLÆRINGSMÆSSIG BIAS - 6
- ALGORITMISK BIAS OG DENS KONSEKVENSER I DEN VIRKELIGE VERDEN - CASE-EKSEMPLER - 16
- VÆRKTØJER TIL AT OPDAGE ALGORITMISK BIAS - 28
- PROJEKT | AI I REKRUTTERING - 35

INDLEDNING



BILLEDKILDE | Freepik

I DENNE KOMPETENCEENHED FINDER DU FØLGENDE EMNER:

- Casestudier, der præsenterer forskellige algoritme- og maskinlæringsbias, hvori TikTok og ansigtsgenkendelsesteknologier indgår.
- En oversigt over værktøjer, der er designet til at identificere algoritmisk bias og dens konsekvenser.
- Et dybdegående projekt om brugen af AI i rekruttering, der fremhæver bias i forskellige projektfaser.
- Rollen for fairness-bevidste tilgange til maskinlæring og implementeringen af bias-revisionsprocesser.

VED AFSLUTNINGEN AF KOMPETENCEENHEDEN, BØR DU VÆRE I STAND TIL AT:

- Identificere forskellige typer og årsager til algoritmisk bias og eksempler på typer af bias i maskinlæringsdatasæt.
- Evaluere algoritmisk bias og dens konsekvenser i den virkelige verden.
- Analysere forskellige værktøjer til at opdage algoritmisk bias.
- Analysere et projekts faser – Case: "AI i rekruttering"

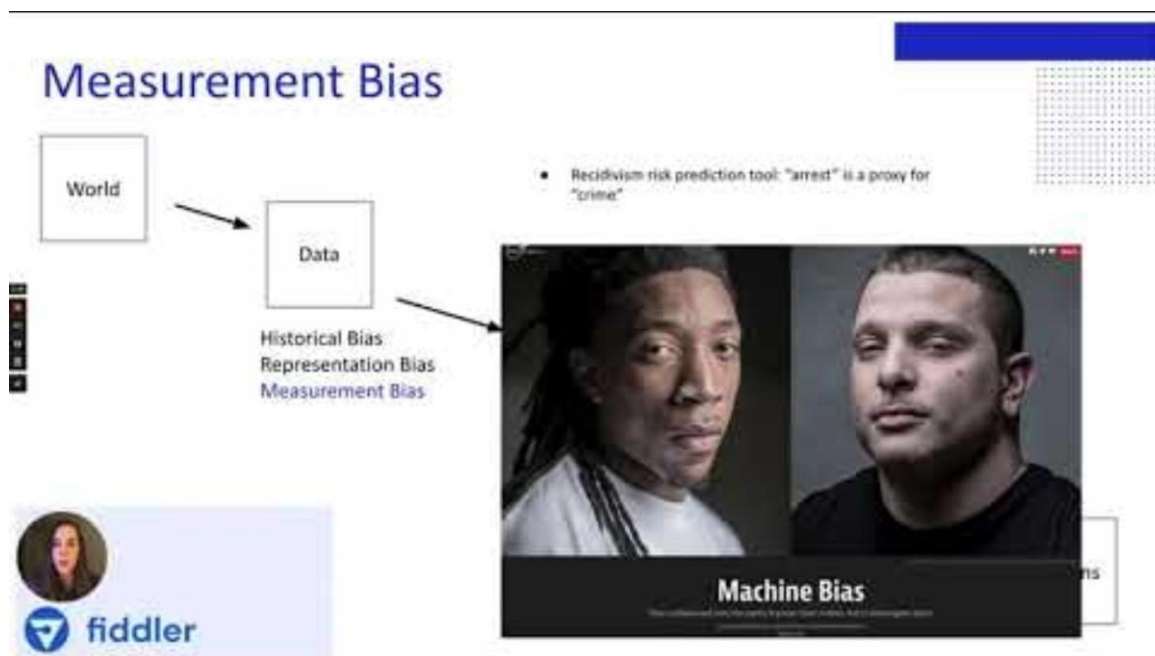
ALGORITMISK OG MASKINLÆRINGSMÆSSIG BIAS



BILLEDKILDE | Freepik

Algoritmisk og maskinlæringsmæssig bias

HVAD ER ALGORITMISK BIAS OG HVORFOR ER DET VIGTIGT AT OVERVEJE?



BILLEDKILDE | MANGLENDE INFO

- I maskinlæring bruger algoritmer datasæt eller træningsdata til at lære mønstre.
- Algoritmisk bias opstår, når algoritmen træffer beslutninger, der systematisk forfordeler eller diskriminerer visse grupper af mennesker, f.eks. på grund af race eller køn.
- Algoritmisk bias kan opstå, hvis en bestemt gruppe mennesker er underrepræsenteret i dataene, eller hvis eksisterende samfundsmæssige bias er indlejret i selve dataene.
- Det er afgørende at forstå årsagerne til og konsekvenserne af algoritmisk bias for at kunne udvikle etiske og troværdige AI-løsninger.

(Harvard Business Review, S. Friis & J. Riley, 2023)

Algoritmisk og maskinlæringsmæssig bias

Forskellige typer af algoritmisk bias

Algoritmiske bias kan f.eks. inddeles i tre kategorier:



1. **Allerede eksisterende bias.** Fordomme eller bias i et system, der eksisterer uafhængigt og som regel før systemet blev skabt. De kan komme ind i et system med vilje eller utilsigtet, enten gennem bevidste handlinger eller ubevidste handlinger, nogle gange endda med gode intentioner.



2. **Teknisk bias** opstår på grund af tekniske begrænsninger eller tekniske overvejelser. Kilder til teknisk bias kan findes i flere etaper af designprocessen.

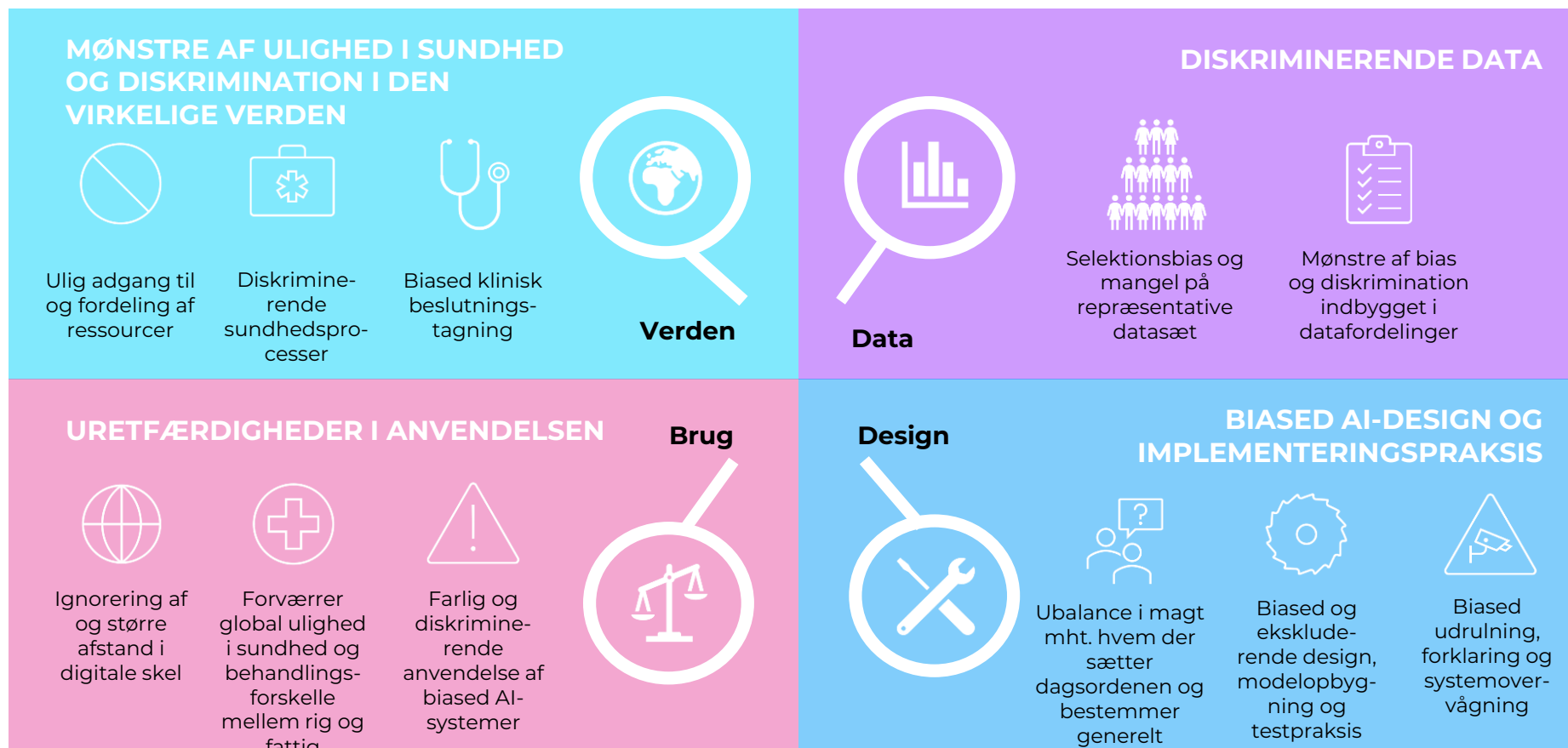


3. **Emergent bias** eller 'nyfremkomne bias' opstår i en brugskontekst med rigtige brugere. Denne type bias opstår typisk, efter at designprocessen er afsluttet, ofte på grund af ændringer i samfundets viden, befolkning eller kulturelle værdier.

Friedman og Nissenbaum (1996)

Algoritmisk og maskinlæringsmæssig bias

Forskellige årsager til bias



BILLEDKILDE | Billedet er redesignet (oprindeligt billede: British Medical Journal)

Digitale teknologier har en tendens til at forstærke uligheder, som f.eks. kan vise sig som diskriminerende applikationer. Dette er en årsag til diskrimination i AI's beslutningstagning, som skyldes etiske 'huller' i udviklingen, implementeringen eller tilgængeligheden af disse teknologier.

Algoritmisk og maskinlæringsmæssig bias

Eksempel på typer af bias i datasæt til maskinlæring

Udviklere træner maskinlæringsmodeller ved at fodre dem med træningseksempler. Når mennesker er involveret i at kuratere og levere data, kan det forårsage bias.

Lad os analysere følgende typer af bias:

- Bias i rapporteringen
- Gruppetilskrivningsbias
- Implicit bias
- Automatiseringsbias
- Bias i udvælgelsen

Algoritmisk og maskinlæringsmæssig bias

Eksempel på typer af bias i datasæt til maskinlæring

Bias i rapporteringen

- Hyppigheden af forekomster, egenskaber eller resultater i et datasæt afspejler ikke forekomsten i den virkelige verden.
- Bias opstår, når man fokuserer på at dokumentere usædvanlige eller mindeværdige omstændigheder.
- Almindelige situationer kan blive overset, da de antages at være selvindlysende.

Eksempel

I træningen af en sentiment-analysemodel (meningsanalysemodel) forudsiger den, om boganmeldelser er positive eller negative baseret på brugerindlæg på et populært websted. De fleste anmeldelser i træningsdataene udtrykker ekstreme synspunkter (at man enten elsker eller hader en bog), fordi folk er mindre tilbøjelige til at anmelde en bog, hvis de har en neutral holdning til den. Derfor har modellen svært ved at forudsige folks meninger korrekt i anmeldelser, der bruger et subtilt sprog til at beskrive bogen.

Algoritmisk og maskinlæringsmæssig bias

Eksempel på typer af bias i datasæt til maskinlæring

Gruppetilskrivningsbias

Gruppetilskrivningsbias er tendensen til at anvende individuelle karakteristika på en hel gruppe, som man tilhører. Der er to hovedformer:

- **In-group bias/bias i gruppen:** at foretrække medlemmer af sin egen gruppe eller medlemmer med fælles karakteristika.

Eksempel | To IT-udviklere, der træner en model til CV-screening af softwareudviklere, foretrækker måske ansøgere fra det samme IT-studie, som de selv har gået på, fordi de antager, at de ansøgere er mere kvalificerede.

- **Homogenitetsbias uden for gruppen:** Stereotypisering af personer i en gruppe, du ikke tilhører, idet du betragter deres karakteristika som mere ensartede.

Eksempel | To IT-udviklere, der træner en model til CV-screening, antager måske, at alle ansøgere, der ikke har gået på et bestemt IT-studie, mangler ekspertise til at kunne udføre jobbet.

Algoritmisk og maskinlæringsmæssig bias

Eksempel på typer af bias i datasæt til maskinlæring

Implicit bias

Implicitte bias opstår, når der gøres antagelser baseret på personlige mentale modeller og erfaringer, som måske ikke gælder universelt.

Eksempel | En IT-udvikler, der træner en model til genkendelse af gestik, bruger en hovedrysten til at indikere "nej". Men nogle steder betyder en hovedrysten "ja".

- En almindelig type implicit bias er **bekræftelsesbias**, hvor de, der bygger modellerne, ubevidst behandler data for at bekræfte eksisterende overbevisninger.
- Nogle gange kan de, der bygger modellerne, blive ved med at træne dem, indtil resultaterne stemmer overens med deres oprindelige hypotese, kendt som **forskerbias eller eksperimentatorbias**.
- **Eksempel** | En IT-tekniker, der skaber en model for hundeaggressivitet, associerer hyperaktivitet med et møde med en lille puddelhund, dengang vedkommende var barn. Selvom modellen oprindeligt skildrer puddelhunde som rolige, fortsætter teknikeren med at omskole modellen, indtil den viser, at mindre pudler er mere aggressive, hvilket bekræfter teknikerenes oprindelige bias.

Algoritmisk og maskinlæringsmæssig bias

Eksempel på typer af bias i datasæt til maskinlæring

Automatiseringsbias

Præference for resultater fra automatiserede systemer frem for ikke-automatiserede, uanset deres respektive fejlprocenter.

Eksempel

Nogle softwareingeniører hos en tandhjulsproducent var begejstrede over at kunne introducere en nyligt trænet "innovativ" model, der var designet til at opdage defekte tandhjul. Men deres begejstring forsvandt, da en fabrikschef påpegede, at modellens præcision og tilbagekaldelsesprocent var 15 % lavere end de rater, som menneskelige kontrollanter kunne opnå.

Algoritmisk og maskinlæringsmæssig bias

Eksempel på typer af bias i datasæt til maskinlæring

Bias i udvælgelsen

Udvælgelsesbias opstår, når eksempler i et datasæt ikke er repræsentative for fordelingen i den virkelige verden. Det kan vise sig på flere måder:

- **Dækningsbias:** Dataene er ikke repræsentativt udvalgt.

Eksempel | En AI-model er trænet til at anbefale restauranter baseret på kundeanmeldelser. Men modellen er kun trænet på anmeldelser af en populær restaurantkæde og ignorerer anmeldelser af andre restauranter. Derfor afspejler AI-modellens anbefalinger muligvis ikke alle potentielle kunders forskellige smag og præferencer, da den ikke har taget højde for udtalelser fra kunder, der har besøgt andre restauranter.

- ***Non-response bias* eller deltagelsesbias:** Data bliver ikke repræsentative på grund af manglende deltagelse under dataindsamlingen.

Eksempel - Et AI-baseret filmanbefalingssystem indsamler feedback fra brugere af en bestemt streamingplatform, men ignorerer feedback fra brugere af andre platforme, hvilket resulterer i, at deres præferencer ikke er tilstrækkeligt repræsenteret. Som følge heraf afspejler de AI-genererede filmanbefalinger muligvis ikke alle brugeres forskellige smag på grund af den manglende feedback fra brugere af forskellige streamingplatforme.

- **Selektionsbias:** Der anvendes ikke korrekt randomisering under dataindsamlingen.

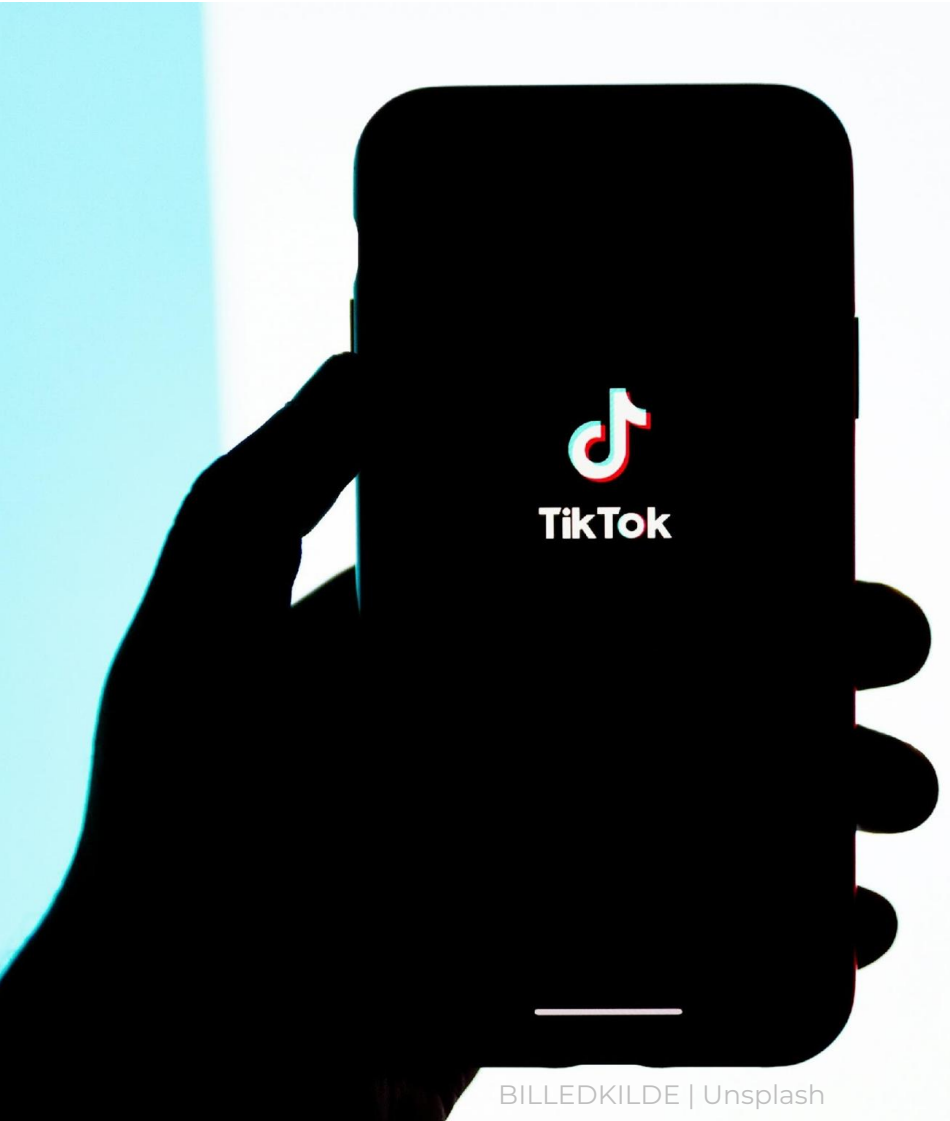
Eksempel | En AI-model trænes til at forudsige det fremtidige salg af et nyt produkt ved at spørge folk, der har købt det, og folk, der har købt et lignende produkt, om deres købsvaner. I stedet for at vælge folk tilfældigt, valgte de de første 200, der svarede på undersøgelsen, hvilket kan betyde, at de var mere begejstrede for produktet end køberne generelt.

ALGORITMISK BIAS OG DENS KONSEKVENSER I DEN VIRKELIGE VERDEN - CASE-EKSEMPLER



BILLEDKILDE | Freepik

CASE - TikToks skadelige algoritme



Finlands nationale medievirksomhed Yle undersøgte, hvilken slags indhold TikToks algoritme ville vise til en **13-årig teenagepige, der led af depression og et forvrænget kropsbillede.**

- Yle skabte en profil for den **fiktive pige, Ella**, som skulle være den deprimerede 13-årige pige.
- Yles datalog surfede på TikTok med Ellas profil i lidt over fem timer.

Se de videoer, som TikTok viste for **Ella i Yles artikel**

CASE - TikToks skadelige algoritme

Resultaterne var alarmerende

1. TIME - skadeligt indhold **7%** ud af 100%.

(Videoer om psykiske problemer og madproblemer blev regnet som skadeligt indhold).

- Først - muntre videoer om mad, hvalpe, katte, skateboarding og vittigheder.
- Ella ser madvideoerne lidt længere.
- Efter kun 5 minutter stopper Ella for at se en video om en kvinde, der skriger.
- Kort tid efter viser TikTok en video om antidepressiva.
- Flere videoer om mental sundhed og lette måltider.

"Tiktoks algoritme lærer hurtigt, hvilke emner en browser er interesseret i."

2. TIME - skadeligt indhold **28%** ud af 100%.

- Tonen i videoerne bliver mørkere.
- Lejlighedsvist indhold om emner som hårpleje og dyr.
- **Indhold relateret til selvmord og selvskade er nu begyndt at optræde i videoerne.**
- Blandt videoerne er der også indhold, som f.eks. opmuntrer folk til at holde ud på trods af vanskeligheder.

Yle, 11. november 2023

CASE - TikToks skadelige algoritme Resultaterne var alarmerende

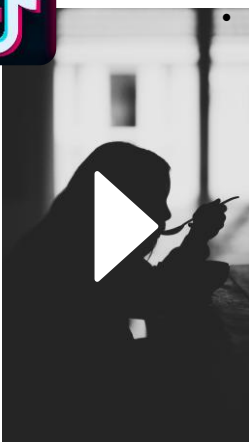
"I en video fortæller en kvinde med en psykisk lidelse, hvordan hun nogle gange har lyst til at tage sit eget liv."

CASE - TikToks skadelige algoritme

Resultaterne var alarmerende

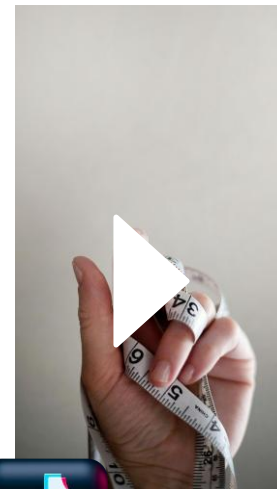
3. TIME - skadeligt indhold 68% ud af 100%.

- Lige før starten på den tredje browsingtime viser Tiktok Ella den første video om spiseforstyrrelser.
- Efter den første video begynder der at dukke mere indhold om spiseforstyrrelser op, og Ella stopper op for at se det.
- Mere indhold om mad med lavt kalorieindhold og selvskade.



4. TIME - skadeligt indhold 57% ud af 100%.

- Lidt færre videoer om depression og selvskade.
- I stedet synes Tiktok-algoritmen at have opfanget, at Ella er særligt interesseret i lette måltider og det giftige indhold, der opfordrer hende til at begrænse sin spisning.
- Flere mad- og væggtabsvideoer end tidligere.

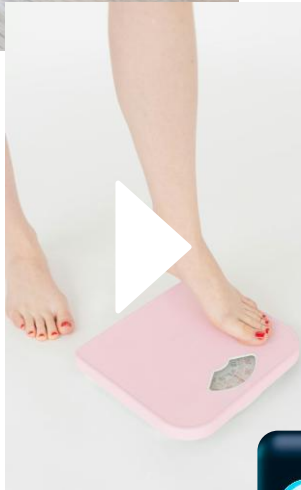
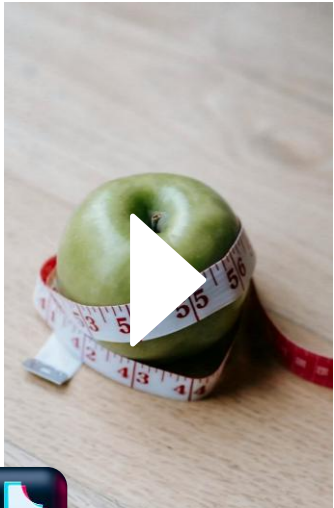


Yle, 11. november 2023

ALGORITMISK BIAS OG DENS KONSEKVENSER I DEN VIRKELIGE VERDEN

CASE - TikToks skadelige algoritme

Resultaterne var alarmerende

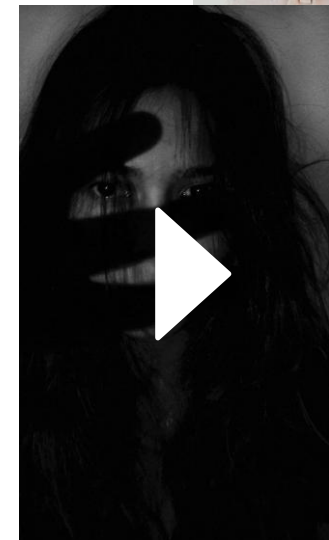
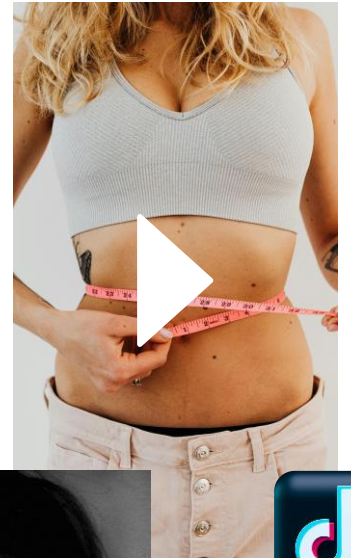


5. TIME - skadeligt indhold **65%** ud af 100%.

- Endnu flere triste videoer dukker op på Ellas skærm.
- Videoer, der forherliger spiseforstyrrelser, og som nu direkte giver konkrete råd om, hvordan man begrænser sin spisning.

VED SLUTNINGEN AF BROWSINGEN - skadeligt indhold **95%** ud af 100%.

- I slutningen af browsingprocessen havde Tiktok-algoritmen vist Ella et stort antal videoer om depression, selvmordstanker, kaloriefattig kost og spiseforstyrrelser.



Yle, 11. november 2023

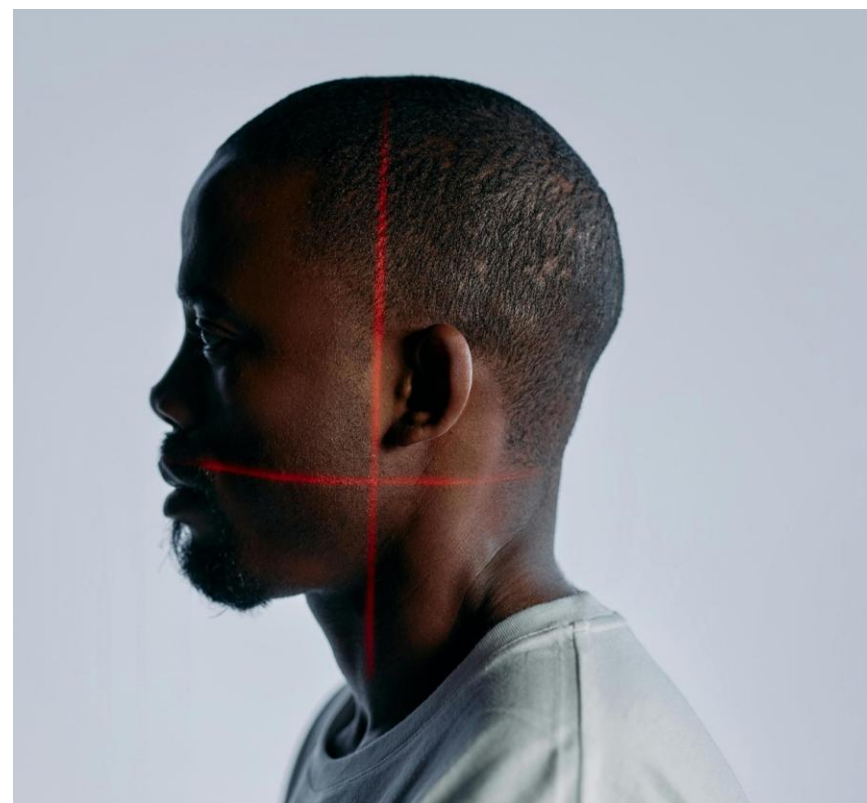
CASE - TikToks skadelige algoritme Resultaterne var alarmerende

Ifølge socialpsykolog Suvi Uski viser algoritmen brugeren indhold, der interesserer dem så meget som muligt, uanset hvor skadeligt det er. Uski siger, at det er særligt problematisk, fordi Tiktok samtidig er langt den mest vanedannende af de sociale apps.

Etiske bekymringer i forbindelse med ansigtsgenkendelse

Ansigtsgenkendelse - hvad og hvordan?

- Et ansigtsgenkendelsessystem **scanner og identificerer en persons ansigt ved at sammenligne det med et lagret digitalt billede eller et billede fra et videoklip** i en database ved hjælp af en AI-algoritme.
- Teknologien fungerer **ved at identificere ansigtstræk** i et billede **og sammenligne dem med andre** billeder eller videoer i databasen ved hjælp af en AI-algoritme.



BILLEDKILDE | Pexels

Ansigtsgenkendelse - eksempler på brug



1 **Sundhed.** Automatisk scanning af en patients ansigt ved hjælp af ansigtsgenkendelse for at fremsøge sygehistorie og forsikringsoplysninger. Bruges også til at diagnosticere medicinske lidelser.



2 **Detailhandel.** Amazon bruger ansigtsgenkendelse til kontaktløse betalinger. Kan bruges til at personliggøre indkøbsoplevelsen, f.eks. i forbindelse med kundeloyalitetstilbud. Kan også bruges til at forbedre sikkerheden i detailhandlen.



3 **Bank og finans.** Bruges til kundeidentifikation og -verifikation. For at fremskynde kundens onboarding-proces. Og for at mindske røverier og banksvindel.

4 **Retshåndhævelse og sikkerhed.** Til at identificere og registrere kriminelle og personer af som man måske finder mistænkelige. At opdage flygtninge ved hjælp af overvågningskameraer i en by, f.eks. ved vejkryds, tog- og busstationer og lufthavne.

ALGORITMISK BIAS OG DENS KONSEKVENSER I DEN VIRKELIGE VERDEN

CASE - Uetisk indsamling af ansigtsgenkendelsesdata



Video om ansigtsgenkendelse i lufthavne på Washington Posts Youtube-kanal

- ▶ Se videoen om ansigtsgenkendelse i lufthavne.
- ? Hvilke etiske spørgsmål opstår i løbet af videoen?

Washington Post, 10. juni 2019

Etiske betænkeligheder ved indsamling af ansigtsgenkendelsesdata

"FBI, ICE finder ud af, at kørekortbilleder er en guldgrube for ansigtsgenkendelsessøgninger"

[Læs nyhedsartiklen på The Washington Post den 7. juli 2019](#)

Agenter fra Federal Bureau of Investigation og Immigration and Customs Enforcement har brugt statslige kørekortdatabaser som en rig kilde til ansigtsgenkendelse og scannet gennem millioner af amerikaneres billeder uden deres viden eller samtykke.

Drew Harwell, The Washington Post, 7. juli 2019

Racemæssig bias i ansigtsgenkendelsessystemer

"Asiatiske og afroamerikanske personer havde op til 100 gange større risiko for at blive fejlidentificeret end hvide mænd, afhængigt af den pågældende algoritme og type søgning."

[Læs nyhedsartiklen på The Washington Post den 19. december 2019](#)

En føderal undersøgelse bekræftede, at **mange af ansigtsgenkendelsessystemerne har racemæssige bias.**

Drew Harwell, The Washington Post, 19. december 2019

VÆRKTØJER TIL AT OPDAGE ALGORITMISK BIAS



BILLEDKILDE | Freepik

Værktøjer til at opdage algoritmisk bias

UNESCO's Ethical Impact Assessment

- Udviklet af UNESCO, 2023
- Et værktøj til at foretage en etisk vurdering ved at evaluere og identificere AI-systemers fordele og risici.
- Værktøjet består af forskellige spørgsmål, der er klar til at blive udfyldt
- Både åbne og multiple choice-spørgsmål
- Etikken er tænkt ind i hele vurderingen
- Starter med afgrænsningsspørgsmål efterfulgt af spørgsmål relateret til UNESCO's principper.

Værktøjet hjælper dig med at besvare spørgsmål som f.eks.:

- Hvem skal interagere med AI-løsningen?
- Hvad er rollerne og ansvarsområderne i AI-teamet?
- Hvem er de interessenter, der vil blive påvirket af AI-systemet?
- Hvordan er der tænkt på sikkerhed og sikkerhed?
- Eventuelle spørgsmål vedrørende bæredygtighed, privatlivets fred, gennemsigtighed, forklarlighed og ansvarlighed

[Find det her](#)

Udviklet af UNESCO, 2023

Figure 2: Positionality Matrix (developed by The Alan Turing Institute)



UNESCO, 2023

BILLEDKILDE UNESCO

Værktøjer til at opdage algoritmisk bias

UNESCO's Ethical Impact Assessment

Hvordan opdager værktøjet bias?

- Afsnittet 'Fairness, ikke-diskrimination, mangfoldighed' fokuserer mere detaljeret på bias, der optræder i data
- Bias håndteres også ved at være opmærksom på roller og ansvar.

6.2. Procedural Assessment

Throughout this section, when responding to questions about testing with particular groups, the project team should consider especially – but not only – race, colour, descent, gender, age, language, religion, political opinion, national origin, ethnic origin, social origin, economic or social condition of birth, and disability. Please specify if testing was conducted on groups that combine several of these criteria i.e. if the system has been tested in terms of intersectionality.

6.2.1. Preventing discriminatory outcomes:

6.2.1.1. Has the algorithm been tested with different groups?

6.2.1.1.1. Was there a difference in terms of accuracy rate (or any other performance metric used)? Please describe any difference of this kind.

6.2.1.1.2. Was there a discriminatory effect for particular groups?

6.2.2. Data quality and preventing discriminatory bias:

6.2.2.1. Are processes in place to test data against biases?

6.2.2.1.1. Have you undertaken an analysis of the data to prevent societal and historical biases in data?

6.2.2.1.2. Is the data well-balanced and does it reflect the diversity of the targeted end-user population?

6.2.2.1.3. Are there any differences you can foresee between the data used for training and the data processed by the AI system which could result in the AI system producing discriminatory outcomes or performing differentially for different groups?

6.2.2.1.4. Have you developed a process to document how data quality issues can be resolved during the design process?

6.2.2.1.5. Did you put in place educational and awareness initiatives to help AI designers and developers gain awareness of the possible bias they can introduce through the design and development of the AI system?

6.2.2.1. Are processes in place to test data against biases?

Værktøjer til at opdage algoritmisk bias

ALTAI

- Liste til at vurdere troværdighed i AI
- Udviklet af Europa-Kommissionens Ekspertgruppe på højt niveau (AI HLEG) om kunstig intelligens
- 'The Assessment List for Trustworthy AI (ALTAI)'
- Liste med 7 krav til troværdig AI, der er defineret mere specifikt på hjemmesiden
- Indeholder en anbefaling om, hvordan man udfylder ALTAI, og hvem der skal involveres.
- [Find den her](#)

Sections of the ALTAI

- 📁 Human Agency and Oversight
- 📁 Technical Robustness and Safety
- 📁 Privacy and Data Governance
- 📁 Transparency
- 📁 Diversity, Non-Discrimination and Fairness
- 📁 Societal and Environmental Well-being
- 📁 Accountability

Is the **AI system** designed to interact, guide or take decisions by human **end-users** that affect humans ('**subjects**') or society? ⓘ *

- ☐ Yes
- ☐ To some extent
- ☐ No
- ☐ Don't know

Did you put in place procedures to avoid that **end-users** over-rely on the **AI system**? ⓘ *

- ☐ Yes
- ☐ No

Værktøjer til at opdage algoritmisk bias

ALTAI

Hvordan opdager værktøjet bias?

- ALTAI-værktøjets afsnit 'Mangfoldighed, ikke-diskrimination og retfærdighed' fokuserer på at undgå uretfærdig bias.
- ALTAI-værktøjet vurderer algoritmisk bias gennem målrettede spørgsmål og evaluerer inputdata og algoritmedesign for bias, samtidig med at det fremmer inklusion, tilgængelighed og interessenters deltagelse for at sikre retfærdighed og gennemsigtighed i AI-systemer.

Avoidance of unfair bias

Did you establish a strategy or a set of procedures to avoid creating or reinforcing **unfair bias** in the **AI system**, both regarding the use of input data as well as for the algorithm design?  *

- ☐ Yes
☐ No

Did you consider a mechanism to include the participation of the widest range of possible stakeholders in the **AI system**'s design and development?  *

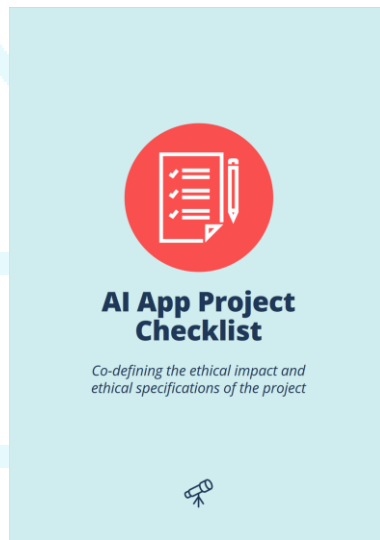
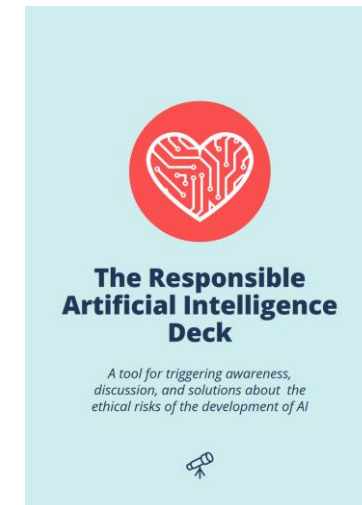
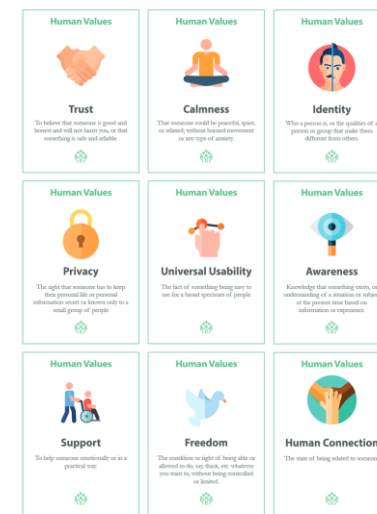
- ☐ Yes
☐ No

Værktøjer til at opdage algoritmisk bias

Ethical Toolkit for the Development of AI Applications

- Udviklet af Mario Sosa Hidalgo, (TU Delft Industrial Design Engineering)
- Undersøger etikken i udviklingen af AI-applikationer
- Innovativ værktøjskasse med en aktiv, konkret og visuel tilgang

[Find den her](#)

The image shows the 'AI APP PROJECT CHECKLIST' form. It is a structured form with sections for: 1. Goal, 2. Team Accountable for App, 3. Context of the Project, 4. Stakeholders Impacted, 5. Type of Impact, 6. AI Algorithm & Learning Model used, 7. Data Type, 8. Data Source, and 9. Ethical Principles Checklist. The checklist includes categories like Data Privacy, Data Safety, Explainability, Transparency, Accountability, Fairness, Value Alignment, Human Well-Being, Governance, and User Safety.

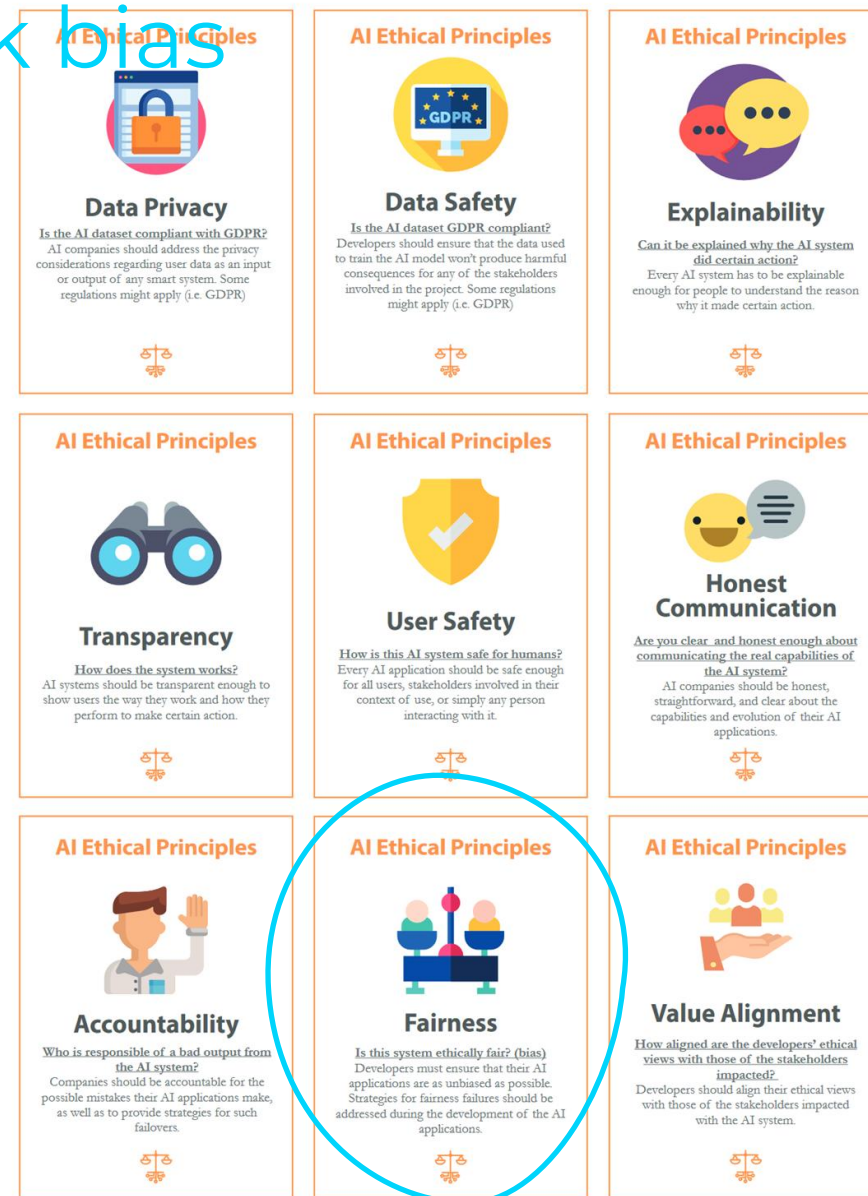
Mario Sosa Hidalgo, 2019

Værktøjer til at opdage algoritmisk bias

Ethical Toolkit for the Development of AI Applications

Hvordan opdager værktøjet bias?

Mario Sosas værktøj undersøger bias og anbefaler udviklere at sørge for, at det udviklede AI-system er så upartisk som muligt, og minder udviklerne om strategier for at undgå fairnessfejl.



Top icons by Flaticon.com

Mario Sosa Hidalgo, 2019

PROJEKT | AI I REKRUTTERING



BILLEDKILDE | Freepik

Projekt | AI i rekruttering



PROBLEMFORMULERING OG FORRETNINGSFORSTÅELSE



Casestudie | Baggrund

En global industrivirksomhed oplever ineffektivitet og usikkerhedsmomenter i sin rekrutteringsproces, hvilket lægger et betydeligt pres på HR-afdelingens tid og ressourcer.

I forventningen om fremtidig vækst ser virksomheden også et presserende behov for at ansætte en række specialister, herunder ingeniører, marketingfolk, finansfolk, kodere og projektledere.

Den hårde konkurrence om kompetente medarbejdere og de høje omkostninger, der er forbundet med fejlansættelser, understreger vigtigheden af en vellykket rekruttering. Situationen øger presset på HR-teamet, hvis opgave det er at sikre, at rekrutteringsprocessen bliver en succes og forløber gnidningsløst.

Derfor undersøger virksomheden, om der er udviklet en AI-løsning, der er skræddersyet til dens rekrutteringsbehov.

Når man integrerer AI i rekrutteringsprocessen, er det primære mål at forbedre effektiviteten, objektiviteten og den samlede succes af rekrutteringen.

Som et første skridt overvejes fordele og ulemper ved en AI-rekrutteringsløsning.

Projekt | AI i rekruttering

AI-rekruttering - fordele og ulemper

FORDELE

- **Effektivitet og tidsstyring:** AI strømliner opgaver, hvilket kan hjælpe rekrutteringsmedarbejdere til at fokusere på strategi, hvilket fremskynder ansættelse
- **Forbedret kvalitet af ansøgerbasen:** AI analyserer data for at vælge de bedst egnede kandidater, hvilket reducerer medarbejderomsætningen
- **Forbedret ansøgeroplevelse:** AI giver rettidige opdateringer og forbedrer kommunikationen
- **Reduktion af omkostninger:** AI reducerer rekrutteringsomkostningerne ved at reducere det manuelle arbejde med rutineopgaver
- **Øget retfærdighed:** AI reducerer bias og øger objektiviteten i ansættelser -> Forbedring af mangfoldighedsaspektet
- **Grundig analyse af sociale medier:** AI gennemgår sociale medier for at se, hvordan ansøgere passer ind i virksomhedskulturen
- **Global rekruttering:** AI kan overvinde sprogbarrierer, hvilket giver adgang til større talentpuljer

ULEMPER

- **Mindre personligt engagement:** AI gør rekruttering mere mekanisk og afskrækker muligvis kandidater
- **Risiko for bias:** Forkert AI-design eller biased data kan kompromittere mangfoldighed og retfærdighed og fastholde sociale uligheder.
- **Datasikkerhed:** AI's håndtering af persondata øger risikoen for privatlivets fred og cyberangreb
- **Høje investeringsomkostninger:** AI i rekruttering kræver betydelige økonomiske og ressourcemæssige investeringer
- **Menneskelig dømmekraft:** AI-afhængighed kan gøre, at firmaet går glip af dygtige kandidater på grund af dataproblemer
- **Overholdelse af lovgivningen:** Rekrutterings-AI skal tilpasse sig skiftende love for at opretholde retfærdighed
- **Kulturel vurdering:** AI kan måske ikkeeffektivt vurdere evner til kulturel tilpasning eller teamwork

(www.recruiter.com, 2023; Bursell & Roumbanis, 2024; Arivu Recruitment, 2023; Köchling & Wehner, 2020; Albaroudi et al., 2024; Sheard, 2022; Albassam, 2023).

Projekt | AI i rekruttering



PROBLEMFORMULERING OG FORRETNINGSFORSTÅELSE

Forståelse af problemet

Virksomheden begynder med at lave baggrundsundersøgelser for at udvikle en AI-løsning og danner derfor et internt team, der skal identificere de vigtigste udfordringer i den eksisterende rekrutteringsproces. Teamet gennemfører interviews med en række interessenter, herunder HR-afdelingen, nyansatte medarbejdere og ledere, der for nylig har ansat nye teammedlemmer. Teamet indsamler også data om de nuværende rekrutteringsomkostninger og analyserer succesraten for disse tiltag.

Det er afgørende for virksomheden at forstå potentielle medarbejderes opfattelse af brugen af AI i rekrutteringen. Den ønsker at undgå skader på sit omdømme eller brand. For at opnå dette er virksomheden forpligtet til at sikre, at indførelsen af kunstig intelligens ikke kun får bred accept, men også overholder etiske og juridiske standarder.



Projekt | AI i rekruttering



PROBLEMFORMULERING OG FORRETNINGSFORSTÅELSE



Forståelse af systemets brugere

En af de interviewede personer er Lead Recruiter Maria, som har flere års erfaring i virksomheden. Hun har med egne øjne set den stadige ophobning af rekrutteringsopgaver og det voksende pres, det lægger på HR-afdelingen, og som intensiveres, når man ansætter internationale talenter.

Maria er åben og interesseret i mulighederne for at bruge kunstig intelligens til at hjælpe med rekrutteringsopgaver, selv om hun ikke har et indgående kendskab til teknologien i kunstig intelligens.



VÆRKTØJ

Personas er detaljerede brugerprofiler i designet, der repræsenterer de vigtigste træk ved målgrupperne og undgår stereotyper. De indeholder information om adfærd, behov og demografi for at fremme empati og vejlede designteams i at skabe brugercentrerede løsninger. De er nyttige til at se på tværs af konventionelle markedssegmenter for at opnå mere inkluderende designresultater (Kumar, 2013; Stickdorn et al., 2018a, 2018b).

Projekt | AI i rekruttering

Maria - LEAD RECRUITER

#motivationbestemmer
#goodtalents



BILLEDKILDE | Billede lavet af VAMK

Empatisk

Åbent sind

Udgående

Holdspiller

Perfektionist

DEMOGRAFI

Alder: 37 år
Bopæl: Helsinki, Finland
Nationalitet: Guatemalansk
Uddannelse: MBA Human
Resource Management
Job: Ledende
rekrutteringsmedarbejder
Familie: Bor alene med sin kat

ARBEJDSMÅL

At finde nye talenter
Medarbejdertilfredshed
Hurtig
rekrutteringsproces

HOBBIER



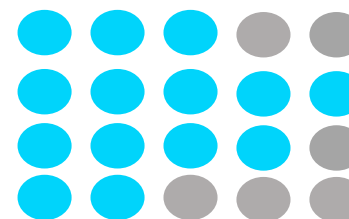
Svømning
Klatring
Læsning

FRUSTRATIONER

- ✓ Høj arbejdsbelastning
- ✓ At finde et godt talent, som takker nej lige før jobstart
- ✓ For langsom rekrutteringsproces får kandidater til at søge andre steder hen
- ✓ Ikke nok tid til at holde kandidaterne informeret under rekrutteringsprocessen
- ✓ Føler sig presset til at finde de perfekte kandidater

FÆRDIGHEDER

Sociale medier
HR-software
Interview
Benchmarking



Maria har store forhåbninger, men hun er også bekymret for, hvilken indflydelse AI-drevet rekruttering vil få på hendes arbejde. Hun er bekymret for, hvordan indførelsen af AI kan ændre HR-medarbejdernes funktioner i fremtiden: Hvilke nye evner og kompetencer skal hun have? Vil hendes bidrag blive mindre værdsat? Er der stadig brug for hendes ekspertise? Og i hvilket omfang kan rekrutteringsprocessen automatiseres?

Projekt | AI i rekruttering



PROBLEMFORMULERING OG FORRETNINGSFORSTÅELSE



Forstå rekrutteringsrejsen

For at få en omfattende forståelse af rekrutteringsprocessen udvikler virksomheden en rekrutteringsrejse, hvor Maria og rekrutteringsteamet indgår. I samarbejde med linjeledere navigerer de gennem hvert trin på rejsen for at identificere væsentlige frustrationer og tidskrævende trin. Derudover undersøger de den potentielle integration af AI på hvert trin for at forbedre rekrutteringsrejsens effektivitet.



VÆRKTØJ

Kort over rejsen er et designværktøj, der skitserer den trinvis oplevelse, som en bruger har med en service, herunder interaktionsstadier, touchpoints og eventuelle udfordringer. Det indfanger også følelsesmæssige reaktioner for at fremhæve kvaliteten af oplevelsen. Kort over rejsen kan repræsentere aktuelle eller fremtidige interaktioner og spænder fra brede, generelle oversigter til detaljerede øjebliksbilleder (Kumar, 2013; Stickdorn et al., 2018a).

Projekt | AI i rekruttering

Maria's - rekrutteringsrejse



IDENTIFICERING AF ANSÆTTELSESBEHOV

Behovet for en ny person identificeres af linjelederen og Maria informeres



TALENTSØGNING

Maria har brug for at identificere og tiltrække nye talenter



INTERVIEWS

Maria og nærmeste leder interviewer gode kandidater



INTRODUKTION OG INDLEDNING

Maria sender induktionsmateriale til medarbejderen, og introduktionen fortsætter, når medarbejderen ankommer



UDARBEJDELSE AF JOBBESKRIVELSEN

Jobbeskrivelse udarbejdes for at tydeliggøre de specifikke færdigheder, der er brug for



SCREENING OG UDVÆLGELSE

Maria har brug for at identificere de rigtige ansøgere fra en stor pulje af kandidater



EVALUERING OG ANSÆTTELSESTILBUD

Udvælgelse og baggrundstjek med reference. Ansættelsestilbuddet sendes til kandidaten.

Hvordan AI kan bruges i forskellige faser



IDENTIFICERING AF ANSÆTTELSESBEHOV

- Analyse af eksterne faktorer som markedskrav, konkurrenters handlinger og HR-målinger
- Vurdering af historiske arbejdstendenser, demografi og kompetencekrav sammen med interne evalueringer
- Sporing af parametre som medarbejderomsætningshastighed og tiden der går med besættelse af stillinger med henblik på effektiv rekruttering



UDARBEJDELSE AF JOBESKRIVELSEN

- Undgå menneskelig bias i udarbejdelsen af jobbeskrivelsen
- Justering af sprog i reklamer og overvågning af indvirkningen på ansøgningsvolumen og -kvalitet



TALENTSØGNING

- Muliggør skræddersyet kommunikation på digitale platforme og sociale netværk
- AI-bots, der lokaliserer både aktive og passive jobsøgere på tværs af platforme og tidligere ansøgerdatabaser
- Brug af tekstmining-strategier til at forbedre jobannoncens appel med sætninger, der tiltrækker kandidater



SCREENING OG UDVÆLGELSE

- Indledende screening ved at analysere semantik og forudsige langsigtet ansættelsespotential og negativ adfærd
- Videointerviews hvor verbale og nonverbale signaler kan vurderes for at evaluere personlighed og kognitive færdigheder
- Analyse af sociale medier ved at undersøge onlineadfærd for at vurdere tilpasningen til virksomhedskulturen
- CV-analyse med NLP-teknikker, der kan berige CV-evalueringer

Projekt | AI i rekruttering

Hvordan AI kan bruges i forskellige faser



INTERVIEWS

- AI-interviewteknikker bruges mest i pre-screening-fasen



EVALUERING OG JOBTILBUD

- Analyse af kandidatens tidligere stillinger for at skabe et skræddersyet tilbud, som kandidaten sandsynligvis vil acceptere.



INTRODUKTION OG INDSLUSNING

- Automatisering af dokumentoprettelse
- Overvågning af onboarding-forløbet
- Skræddersyede uddannelsesforløb

Recruiter.com; Chen, 2022; Hunkenschroer & Luetge, 2022, Lawton ; Albouradi 2024; Marr 2023; Fritts & Carbera,2021; Vivek, 2023

Projekt | AI i rekruttering



DESIGN-FASE

At forstå personlige bias

Efter en klar problemformulering har virksomheden valgt at gå videre med en AI-drevet løsning. Du er blevet udnævnt til ledende AI-udvikler med ansvar for at føre tilsyn med implementeringen af den valgte AI-teknologi.

Virksomheden forstår vigtigheden af at samle en mangfoldig gruppe med flere interesser for at sikre, at en række forskellige perspektiver indgår, hvorved projektet gerne skulle blive en succes.

I et forsøg på at håndtere potentielle, ubevidste bias i teamet kræver virksomheden, at alle gruppemedlemmer deltager i en øvelse, baseret på Alan Turing, som kaldes "Positionality Matrix".



VÆRKTØJ

Alan Turings "positionality matrix" er designet til at belyse teamets personlige bias og vise, hvordan socioøkonomiske baggrunde, kulturelle kontekster og individuelle livserfaringer kan påvirke ens dømmekraft og beslutningstagning.



Projekt | AI i rekruttering

Alan Turing test | Refleksion over formål, positionalitet og magt



Personlige egenskaber og identifikation med grupper

Hvordan identificerer jeg mig?
Alder, race og etnicitet, handicap, religion, køn, seksualitet, ægteskabelig status, forældres status, sproglig baggrund



Baggrund mht. uddannelse og arbejde

Hvordan er jeg blevet uddannet og skolet?
Skolegang, uddannelsesniveau, muligheder for advancement og faglig udvikling, beskæftigelseshistorik



POSITIONERINGSMATRIX

I hvilket omfang repræsenterer mine personlige karakteristika, gruppeidentifikationer, socioøkonomiske status, uddannelses- og arbejdsbaggrund, teamsammensætning og institutionelle rammer kilder til magt og fordele eller kilder til marginalisering og ulemper? Hvordan påvirker denne positionalitet min (og mit teams) evne til at identificere og forstå de berørte interessenter og de potentielle konsekvenser af mit projekt?



Institutionelle rammer og teamsammensætning

Hvordan ser min institutionelle kontekst og teamsammensætning ud?
Autoritetsstruktur i mit projektteam, bredere politisk ejerskab og magthierarkier i min organisation, niveauer af beslutningsautonomi, muligheder for at give udtryk for bekymringer og indvendinger, teamdiversitet, kultur for inklusion eller eksklusion



Socioøkonomisk status

Hvad er min socioøkonomiske historie?
Socioøkonomisk status under opvæksten, social mobilitet over tid, nuværende status, socioøkonomiske ambitioner

Projekt | AI i rekruttering



DESIGN-FASE

Kort over interessenter

Ud over at skulle tage hensyn til den interne gruppe af interessenter, der deltager i projektet, er virksomheden også nødt til at tage hensyn til andre relaterede interessenter.

Derfor udarbejdede virksomheden et kort over interessenter, inspireret af G.J. Millers definition af kortlægning af interessenter i en AI-udviklingsproces, og dette kort viser udviklings- og brugerinteressenterne samt de eksterne interessenter.



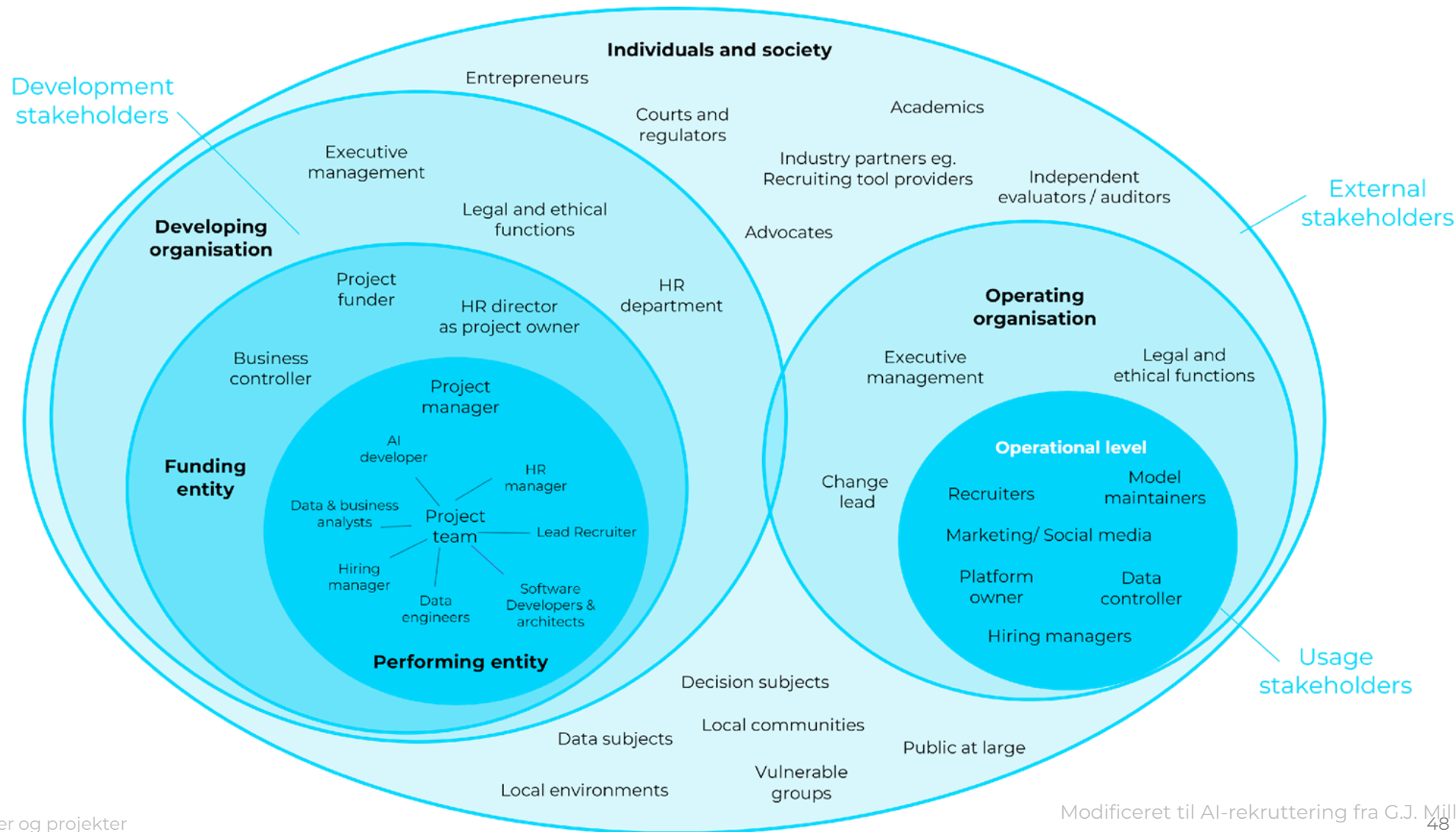
VÆRKTØJ

Kort over interessenter er et visuelt værktøj, der bruges til at identificere og vise rollerne og relationerne mellem alle interessenter, der er involveret i et projekt. Det kan variere fra en simpel kvadrant, der viser niveauer af indflydelse og engagement, til en detaljeret matrix, der beskriver interaktioner mellem interessenter. Kortet over interessenter hjælper med at tydeliggøre interessenternes dynamik og er afgørende for at forstå og styre netværksrelationer (Giordano et al., 2018; SDT, n.d.-b).



Projekt | AI i rekruttering

Kort over interessenter i AI-rekrutteringsprojekt



Projekt | AI i rekruttering



DESIGN-FASE



Matrix til analyse af interessenter

På baggrund af dette kort skal projektlederen inddele interessenterne i grupper for at vide, hvordan de skal holdes informeret, og for at se, hvilken indflydelse de har på projektet.

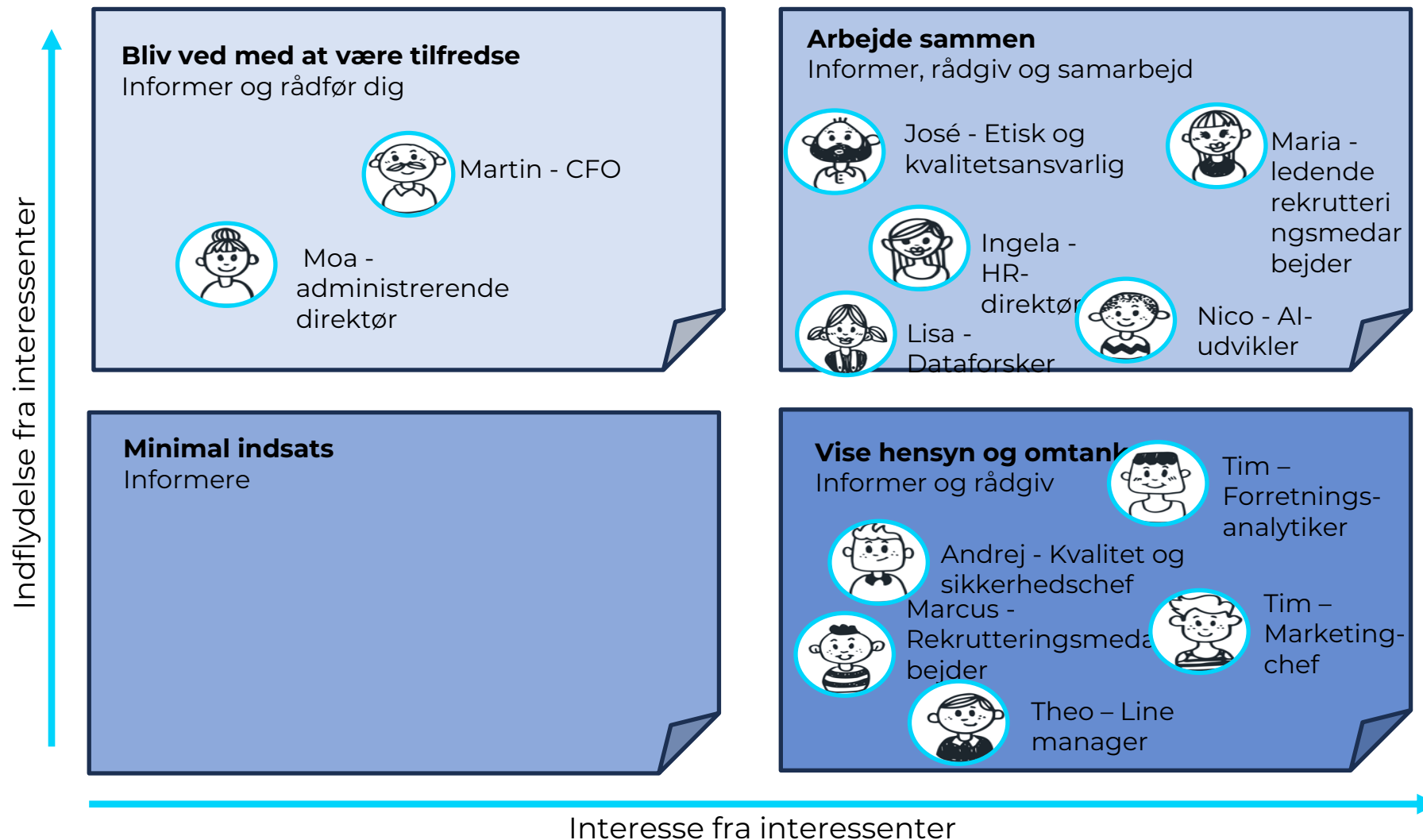


VÆRKTØJ

Matrix til analyse af interessenter indplacerer projektets interessenter efter indflydelse og interesse i et kvadrantnet for at med henblik på at foreslå en engagementstaktik. Den viser, hvordan man kan tilfredsstille interessenter med stor indflydelse og lav interesse, styre dem med stor interesse og stor indflydelse tæt, holde dem, der er interesserede, men mindre indflydelsesrige, informeret og minimalt engagere dem, der ikke har meget af det ene eller det andet (Hoory & Botorff, 2022; Wallbridge, 2023).

Projekt | AI i rekruttering

Matrix til analyse af interessenter



Projekt | AI i rekruttering



DESIGN-FASE



Bias i rekrutteringsfasen - Workshop

Når interessenterne er klarlagt, beslutter virksomheden at afholde en workshop om etik, hvor man gennemgår alle mulige bias i de forskellige ansættelsesfaser.

Interessenter fra området vedr. "Samarbejde" (forrige slide) inviteres til at deltage.

Hvilken slags Post-it-sedler ville du have udfyldt?



VÆRKTØJ

En **workshop** er et kollaborativt designværktøj, hvor en gruppe deltagere indgår i interaktive aktiviteter for at løse et specifikt problem eller arbejde på et projekt. Workshops faciliteres typisk af en leder og kan variere i varighed fra et par timer til flere dage, afhængigt af målene og opgavernes kompleksitet. Denne metode er effektiv til at generere ideer, løse problemer og fremme teamwork og innovation i en gruppe (Wirtz, 2022).

Projekt | AI i rekruttering



Bias i identifikation af ansættelsesbehov

FORANKRINGS-BIAS

Den person, hvis stilling skal besættes, påvirker det forventede ansættelsesbehov

Lavt præsterende person -> Forudsigelse: behov for flere ressourcer

Højt præsterende person -> Forudsigelse: behov for færre ressourcer

STATUS QUO-BIAS

Der kan være modstand mod at ændre rolletypen i organisationen, selvom forretningsbehovene måske siger noget andet.

DEMOGRAFISK SKÆVHED

Forudsigende modeller tager måske ikke højde for demografiske ændringer, som f.eks. en aldrende befolkning.

BIAS I FEEDBACK-LOOP

AI-algoritmen lærer af sine egne tidligere beslutninger, dvs. hvis data om ansættelsesbehov er biased, kan den ende med at skabe større og større forudindtagethed baseret på sine tidlige forudsigelser.

BRANCHE-BIAS

I tilfælde af et for snævert branchesyn kan det bredere fremtidige behov blive overset.

Projekt | AI i rekruttering



Bias i udarbejdelsen af jobbeskrivelsen

AI-skabte jobbeskrivelser genereres for det meste for at **reducere den menneskelige bias** i rekrutteringen -> AI kan identificere sprog, der fører til bias

HISTORISK BIAS

Jobbeskrivelser kan trænes ud fra gamle jobbeskrivelser, som kan være forudindtagede

DEMOGRAFISK BIAS

AI kan foreslå udtryk som "ambitiøs" eller "selvsikker leder" og målrette beskrivelserne til bestemte demografiske grupper.

FORANKRINGS-BIAS

Hvis jobbeskrivelser udarbejdes på baggrund af højtydende medarbejders kernekompetencer, kan jobbeskrivelserne skabe bias.

Projekt | AI i rekruttering



Bias i talentsøgning

Køns-, race- og aldersbias

Forudgående udvælgelseskriterier i jobannoncen, der er valgt af arbejdsgiveren, hvor visse demografiske oplysninger udelades.

Omkostningseffektivitet ved visning af annoncer kan føre til, at jobannoncer vises til dem, der højst sandsynligt kun klikker på **annoncen**, i stedet for til dem, der være dygtige til det pågældende job.

Arbejdsgiverens valg af platform

kan være forudindtaget, da nogle platforme kan være mere brugt af bestemte ansøgergrupper.

Brugen af sociale medier er meget forskellig fra person til person;

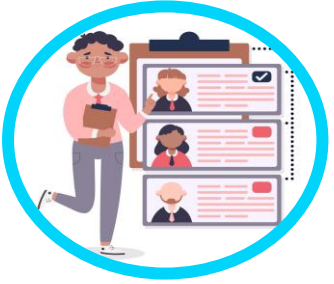
mens nogle måske slet ikke har tid til at bruge f.eks. LinkedIn, giver nogle måske et falsk billede af sig selv, mens andre har flere muligheder for at påvirke deres digitale fodaftryk.

Brugeradfærd forstærker algoritmerne, så de replikerer mønstrene, dvs. hvis en bestemt type hvid mand åbner jobannoncer oftere, vil annoncen være synlig for lignende brugere oftere.

Hvis **rekrutteringsansvarlige kommunikerer mere med kandidater fra bestemte platforme**, kan algoritmerne øge synligheden af annoncer på disse platforme og dermed øge skævvridningen yderligere.

Sheard, 2022; Naakka 2018; Lawton, 2022; Chen 2022; Köchling & Wehner, 2022

Projekt | AI i rekruttering



Bias i screening og shortlisting

Generelt kan AI skabe en mere retfærdig rekrutteringsproces ved at reducere ubevidste fordomme, som mennesker har en tendens til at basere på demografi.

Hvis uddannelsesdata er baseret på en pulje af eksisterende medarbejdere eller en snævert defineret gruppe, kan det føre til utilsigtet diskrimination af underrepræsenterede grupper.

Aggregeringsbias kan opstå, hvis der foretages falske generaliseringer om hele befolkningsgrupper; f.eks. "unge mennesker har tekniske færdigheder" -> lader ældre mennesker ude af betragtning.

Brug af kandidatens digitale fodspor som en del af screeningen, hvis algoritmen er bygget på nøgleord som "energisk" eller "aktiv", kan diskriminere handicappede.

Projekt | AI i rekruttering



Bias i jobinterviews

AI-teknikker til jobinterviews bruges mest i pre-screening-fasen.

De afsluttende samtaler bør ske 'face to face'.

I videoer kan software til følelsesgenkendelse fortolke f.eks. intonationer på forskellige sprog forskelligt.

Nogle kandidater føler sig måske usikre foran et kamera.

Dehumanisering?
Underminering af forholdet mellem medarbejder og arbejdsgiver

Brugen af ansigtsgenkendelsesteknologi kan give anledning til bekymringer om privatlivets fred og fordomme mod visse grupper *.

* Se eksemplet med bias i ansigtsgenkendelse

Sheard, 2022; Naakka 2018; Lawton, 2022; Chen 2022; Köchling & Wehner, 2022

Projekt | AI i rekruttering

Udformning af jobtilbud og onboarding



Det skal overvejes, på hvilket tidspunkt der er behov for interaktion mellem mennesker. Onboarding er at give et billede af virksomheden.



Onboarding

Træningsdata kan være baseret på historiske data og dermed være behæftet med bias.

Onboarding

De personlige fordomme hos den, der har skabt AI-algoritmen, kan påvirke et eventuelt tilbud om uddannelsesforløb.

Værktøjer, der bruges til udformning af jobtilbud, bruger ofte historiske data til at forudse branche-standarder -> kan føre til bias.

Projekt | AI i rekruttering



DATAINDSAMLING OG DATA- FORBEREDELSE



Screening og udvælgelse af kandidater; Definition af datasæt



Mens virksomheden undersøger alle aspekter og muligheder inden for AI-rekruttering, så vurderer den også AI nøje i forhold til screenings- og udvælgelsesfasen. Dette indebærer en dybdegående analyse af CV-scanning og at lave et passende match mellem kandidater og de ledige stillinger.

Ved at anvende AI-værktøjer til at automatisere CV-screeningsprocessen er målet hurtigt og præcist at finde frem til de kandidater, der opfylder jobkravene. Anvendelsen af denne teknologi forventes at føre til en mere strategisk brug af HR-ressourcer, forbedre kvaliteten af de kandidater, der shortlistes, og øge den samlede effektivitet i ansættelsesprocessen ved at afstemme kandidaternes færdigheder og baggrund med virksomhedens krav.

Først skal virksomheden definere de datasæt, som er grundlaget for rekrutteringssystemet, og som algoritmerne henter data fra.

Træning af AI

- Træningsdatasæt spiller en afgørende rolle i udviklingen af algoritmer. Disse datasæt er det fundament, som algoritmerne er bygget på. Formålet med disse datasæt er at fungere som "grundsandheden" eller det faktuelle grundlag, der instruerer algoritmen i, hvordan den skal fortolke og behandle store mængder data.
- De lærer algoritmen at forstå forholdet mellem inputvariabler (den information, der tilføres systemet) og en outputvariabel (den forudsigelse eller beslutning, som systemet træffer på baggrund af input).
- Der opstår en betydelig udfordring, når selve træningsdataene indeholder bias. Disse bias kan afspejle historiske uligheder, fordomme eller simpelthen være resultatet af en ikke-repræsentativ eller ufuldstændig dataindsamling.
- Når en algoritme trænes på forudindtagede data, lærer den disse bias og viderefører dem i sine forudsigelser og beslutninger. Dette fænomen beskrives almindeligvis som "bias in, bias out"-problemet. Det betyder, at hvis inputdataene ("bias in") er forudindtagede, vil maskinlæringsmodellens output ("bias out") også være forudindtaget.
- Det kan føre til uretfærdige, diskriminerende eller fejlbehæftede resultater, hvor kandidater udvælges eller afvises på baggrund af forudindtagede kriterier i stedet for deres sande kvalifikationer eller potentiale.

Projekt | AI i rekruttering

Træning af AI på virksomhedens interne og eksterne data

Interne data

- **Data om medarbejdernes performance og jobhistorik**
 - Data indsamlet internt om medarbejdernes performance og tidligere ansættelser.
- **CV'er indsendt gennem tiden**
 - Gamle CV'er, der er sendt direkte til virksomheden i forbindelse med jobansøgninger.
- **Data om falske negativer**
 - Oplysninger om ansøgere, der fejlagtigt er blevet afvist indsamles fra virksomhedens tidligere rekrutteringsprocesser.
- **Nøgleord og sætninger fra CV'er**
 - Specifikke ord eller sætninger fra CV'er indsamlet gennem virksomhedens ansøgningsproces.
- **Information om uddannelsesmæssig baggrund**
 - Uddannelsesoplysninger indhentet fra ansøgninger til virksomheden.

Eksterne big data

- **Sociale medier**
 - Onlineplatforme, hvor enkeltpersoner deler personlige interesser, erfaringer og karriereresultater. Sociale medier kan give indsigt i en kandidats personlighed, færdigheder og professionelle netværk, hvilket gør sociale medier til en værdifuld kilde til at identificere potentielle talenter.
- **Professionelle hjemmesider**
 - Hjemmesider, der er specielt designet til professionelle netværk og karriereudvikling, såsom LinkedIn. Disse sider indeholder detaljerede profiler, herunder jobhistorik, uddannelse, færdigheder, anbefalinger og arbejdsmæssige resultater, hvilket er en rig kilde til information om talenter.
- **Søgninger på internettet**
 - Brede onlinesøgninger, som virksomheder kan foretage for at finde offentligt tilgængelige oplysninger om en kandidat. Det kan være nyhedsartikler, personlige blogs, publikationer eller andet på nettet, der giver yderligere oplysninger om en persons kvalifikationer, resultater og karakter.

Sheard, 2022; Fernandes, 2021; Ma, 2024

Projekt | AI i rekruttering



MODELUDVIKLING OG TRÆNING



Jobannonce – udvælgelse af nøgleord til algoritmen

Virksomheden har for nylig offentliggjort en jobbeskrivelse og modtaget hundredvis af ansøgninger.

Prototypen til AI-screening og shortlisting scanner ansøgernes CV'er for nøgleord og andre oplysninger, der menes at være relevante for en vellykket ansættelse, f.eks. erfaring, jobtitler, tidligere arbejdsgivere, universiteter og universitetsgrader. Baseret på dette opretter systemet en struktureret ansøgerprofil, og alle kandidater får derefter en score og bliver rangeret.

Jobannonce og relaterede nøgleord



WE ARE LOOKING FOR..

...Communications assistant to be responsible for the creation of content such as media releases, blogs, and social media posts on behalf of our company. You will also be monitoring media and campaign coverage and attending internal and external events.

RESPONSIBILITIES

- Develop and manage diverse content, including media releases, blogs, and social media posts, aligned with the company's strategic goals; monitor media and campaign coverage.
- Support the implementation of internal and external communications strategies and assist in managing the company's image.
- Organize marketing events and provide comprehensive administrative support, including maintaining event calendars and updating contact lists.
- Compile and prepare presentations and reports while tracking project progress and media exposure.

REQUIREMENTS

- Holds a Bachelor's degree in communications, marketing, or related field, with excellent verbal and written communication abilities.
- Proficient in social media strategies and media relations, with a creative and innovative approach.
- Strong organizational skills and attention to detail, capable of multitasking and maintaining positive interpersonal relationships.
- Skilled in using office management and design software, such as Photoshop and InDesign, and knowledgeable in various social media platforms.

(nøgleord oversat til dansk)

"Bachelorgrad i kommunikation"

"Bachelorgrad i marketing" "Skabelse af indhold"

"Medieudgivelser"

"Blogs"

"Indlæg på sociale medier"

"Overvågning af medier" og "kampagnedækning"

"Støtte til implementering af kommunikationsstrategier"

"Håndtering af virksomhedens image"

"Planlægning af markedsføringsevents"

"Yde omfattende administrativ støtte" "Opdatere lister på kontakter"

"Udarbejdelse af præsentationer" og "rapporter" "Holde styr på projektets fremskridt" og "medieeksponering"

"Dygtig til strategier for sociale medier" og "medierelationer"

"Kreativ og innovativ tilgang"

"Stærke organisatoriske evner" og "detaljeorienteret"

"I stand til at multitaske"

"Opretholdelse af positive personlige relationer" "Dygtig til kontorstyring" og "designsoftware"

"Photoshop" og "InDesign"

"Viden om forskellige sociale medieplatforme"

MODELUDVIKLING OG TRÆNING



Sammenligning af to ansøgers CV'er med henblik på formulering af nøgleord

Nu vil vi analysere to ansøgninger for at forstå, hvordan prototypen på AI-applikationen vurderer kandidater i screeningsfasen. Denne analyse vil hjælpe med at identificere, om der er nogen bias, der påvirker udvælgelsesprocessen.

Virksomheden arrangerer en workshop for at se nærmere på AI-drevet CV-screening. Til formålet bruger de et casestudie, baseret på virksomhedens egne aktiviteter.

I løbet af workshoppen analyserer de to CV'er parallelt for at afdække, hvordan forskelle i ordvalg og formuleringer utilsigtet kan føre til, at man overser kvalificerede kandidater.

Dette har til formål at belyse de potentielle bias, der er indlejret i AI-værktøjet, og tilskynder til kritisk tænkning om forbedring af systemet, så det fuldt ud anerkender og værdsætter bredden af en ansøgers kvalifikationer.



JOHN DOE

Communications assistant

EXPERIENCE

Communications assistant

ABC Corporation

2019- Present

- Developed and managed diverse content, including media releases, blogs, and social media posts
- Supported the implementation of comprehensive communications strategies
- Organized multiple marketing events and maintained event calendars
- Compiled presentations and tracked project progress, ensuring alignment with strategic goals

EDUCATION

University of communications

Bachelor's degree in Marketing

2014-2018

SKILLS SUMMARY

- Proficient in Photoshop and InDesign
- Skilled in social media strategies and media relations
- Excellent verbal and written communication abilities
- Strong organizational skills and attention to detail

About Me

A dedicated communications professional with a Bachelor's degree in Marketing, seeking the role of Communications Assistant to leverage extensive experience in content creation, social media management, and event coordination to contribute to the company's strategic goals.



+123-456-7890



hello@doeland.com

123 Anywhere St., Any City



JANE DOE

Digital Content Creator

EDUCATION

CREATIVE UNIVERSITY

BSc in Communications Studies 2018

WORK EXPERIENCE

Digital Content Creator; Creative Media Co
2019 -Present

- Spearheaded content initiatives across digital channels, crafting engaging articles and dynamic social engagements
- Drove strategy for public relations efforts, enhancing brand visibility
- Led the logistics for promotional and networking gatherings, ensuring smooth execution
- Synthesized data into compelling reports and visuals for team updates

COMPETENCIES

- Advanced user of digital design tools and content management systems
- Crafted engaging online dialogue and cultivated media partnerships
- Articulate communicator, both in written formats and oral presentations
- Excelling in project orchestration and meticulous in administrative tasks

PROFILE

Passionate communicator with an academic background in Communications Studies and hands-on experience in crafting engaging narratives and managing digital platforms. Eager to bring my toolkit of creative dissemination and stakeholder engagement to the Communications Assistant position.

CONTACT ME



(123) 456-7890



Jane@doe.com



123 Anywhere St.,
Any city, State,
Country 12345

Fremhævningsbrevet hjælper med at se sammenhængen mellem specifikke formuleringer i CV'erne og nøgleordsbaserede AI-screeningværktøjer.



JOHN DOE

Communications assistant

EXPERIENCE

Communications assistant
ABC Corporation
2019- Present

- Developed and managed diverse content, including **media releases, blogs, and social media posts**
- Supported the implementation of comprehensive communications strategies
- Organized multiple **marketing events** and **maintained event calendars**
- Compiled **presentations** and tracked **project progress**, ensuring alignment with strategic goals

EDUCATION

University of communications
Bachelor's degree in Marketing.
2014-2018

SKILLS SUMMARY

- **Proficient in Photoshop and InDesign**
- Skilled in **social media strategies** and **media relations**
- Excellent verbal and written communication abilities
- **Strong organizational skills** and **attention to detail**

About Me

A dedicated communications professional with a **Bachelor's degree in Marketing**, seeking the role of Communications Assistant to leverage extensive experience in **content creation, social media management**, and event coordination to contribute to the company's strategic goals.

+123-456-7890

hello@doeland.com

123 Anywhere St., Any City



JANE DOE

Digital Content Creator

EDUCATION

CREATIVE UNIVERSITY
BSc in Communications Studies 2018

WORK EXPERIENCE

Digital Content Creator; Creative Media Co
2019 -Present

- Spearheaded **content** initiatives across digital channels, crafting engaging articles and dynamic social engagements
- Drove strategy for public relations efforts, enhancing brand visibility
- Led the logistics for promotional and networking gatherings, ensuring smooth execution
- Synthesized data into compelling **reports** and visuals for team updates

COMPETENCIES

- Advanced user of digital design tools and content management systems
- Crafted engaging online dialogue and cultivated media partnerships
- Articulate communicator, both in written formats and oral presentations
- Excelling in project orchestration and meticulous in administrative tasks

PROFILE

Passionate communicator with an academic background in Communications Studies and hands-on experience in crafting engaging narratives and managing digital platforms. Eager to bring my toolkit of creative dissemination and stakeholder engagement to the Communications Assistant position.

CONTACT ME

(123) 456-7890

Jane@doe.com

123 Anywhere St.,
Any city, State,
Country 12345

Denne tabel illustrerer effekten af forskelle i ordlyd og formuleringer i to CV'er, og hvordan de kan påvirke resultatet af en AI-screeningsproces. Selv om det andet CV har relevante kvalifikationer, kan det blive frasorteret, fordi beskrivelserne ikke stemmer overens med de specifikke nøgleord, som virksomheden har fastlagt for screeningen. Dette viser, hvor vigtigt det er at matche sproget i et CV med de nøgleord, som et AI-system forventer

ASPECT	JOHN DOE	JANE DOE
Jobtitler og terminologi	Bruger standardtermer som "kommunikationsassistent"	Bruger titler som "Skaber af digitalt indhold"
Akademisk baggrund	Nævner udtrykkeligt "bachelorgrad i marketing"	Der står "BSc i kommunikationsstudier"
Beskrivelse af opgaver og færdigheder	Direkte match med søgeord som "Content creation", "Media releases"	Bruger varieret sprog som "skaber engagerende fortællinger", "har stået i spidsen for at skabe indhold"
Tekniske og softwaremæssige færdigheder	Angiver "Photoshop" og "InDesign"	Generel omtale af "digitale designværktøjer og content management-systemer"
Direkte søgeordsmatches	Matcher tæt på mange specificerede nøgleord	Bruger forskellige formuleringer, der måske ikke matcher de specifikke søgeord, der skulle screenes

Projekt | AI i rekruttering



EVALUERING OG AFPRØVNING AF MODELLER



Test af løsningens retfærdighed

En evaluering af algoritmeprototypen afslørede uoverensstemmelser i CV-analysen, hvilket gav anledning til forbedringer for at opnå en mere omfattende forståelse af CV-attributter. Modellen er nu klar til en testfase, inden den tages i brug i den virkelige verden.

Selvom en evaluering af flere aspekter er påkrævet, vil fokus her være på at teste applikationens retfærdighed for at vurdere omfanget af eventuelle bias, der opstår i processen. Fairness-metrikker anvendes til at måle niveauet af bias i applikationens beslutninger.



VÆRKTØJ

IBM's open source-værktøj kan hjælpe med at undersøge, fremfinde og afbøde diskrimination og bias i maskinlæringsmodeller i hele AI-applikationens livscyklus. Se demoen og leg med værktøjet.

[AI-Fairness 360 \(ibm.com\)](https://www.ibm.com/ai-fairness-360)

Projekt | AI i rekruttering

Test af retfærdighed

For at skabe systemer, der er retfærdige for alle og fri for bias såsom køns- eller racediskrimination, er det vigtigt at evaluere løsningens performance på tværs af forskellige befolkningssegmenter.

Der findes forskellige metoder til at vurdere en løsnings retfærdighed, men i denne sammenhæng vil vi udelukkende fokusere på binære klassifikationstests.

Det har vist sig at være en udfordring at definere retfærdighed i matematiske termer, hvilket har ført til manglende konsensus om standardformuleringer. De fleste definitioner af retfærdighed samler sig dog om flere nøglebegreber nedenstående oversættelser i tillempet danske versioner):

- Uvidenhed (*unawareness*)
- Demografisk paritet (*demographic parity*)
- Udlignede odds (*equalized odds*)
- Forudsigende paritet (*predicative rate parity*)
- Individuel retfærdighed (*individual fairness*)
- Kontrafaktisk retfærdighed (*counterfactual fairness*)

Demografisk paritet, forudsigende paritet og udlignede odds grupperes typisk under det overordnede begreb "grupperetfærdighed". Dette vil blive beskrevet nærmere i dette afsnit, især set i lyset af AI i rekruttering.

Projekt | AI i rekruttering

Forvirringsmatrix

En af de binære klassifikationsmetoder er forvekslingsmatrixen.

Forvirringsmatrixen opsummerer forudsigelserne fra en algoritmisk model i forhold til de faktiske resultater, eller grundsandheder, som den blev trænet på. Den viser antallet af både korrekte og forkerte forudsigelser, hvilket giver et klart billede af modellens performance og dens forklarlige **nøjagtighed**.

Forvirringsmatrix i rekrutteringseksemplet

	Ukvalificeret	Kvalificeret
Afvis	Ægte negativ (TN) = afviste ukvalificerede kandidater	Falsk negativ (FN) = afviste, kvalificerede kandidater
Ansæt	Falsk positiv (FP) = ansatte, ukvalificerede kandidater	Ægte positiv (TP) = ansatte, kvalificerede kandidater

(Wortman Vaughan & Wallace, 2022; Zhong, 2018, Saplicki, 2022)

Projekt | AI i rekruttering

Nøjagtighed

I rekrutteringseksemplet illustrerer vi vurderingen af retfærdighed ved at bruge køn som den demografiske variabel, idet vi antager, at der er 100 kvindelige og mandlige ansøgere.

MÆND	Ukvalificeret	Kvalificeret
Afvis	15 (TN)	5 (FN)
Ansæt	20 (FP)	60 (TP)

KVINDER	Ukvalificeret	Kvalificeret
Afvis	60 (TN)	20 (FN)
Ansæt	5 (FP)	15 (TP)

$$\text{Nøjagtighed} = \frac{TP + TN}{TP + FP + TN + FN}$$

I kategorierne for både mandlige og kvindelige er 75 ansøgere klassificeret korrekt.

-> dvs. **nøjagtigheden er den samme for begge demografiske forhold.**

Projekt | AI i rekruttering

Demografisk paritet

Med udgangspunkt i dataene i forvirringsmatrixen er demografisk paritet et kriterium for retfærdighed, hvor sandsynligheden for et gunstigt resultat - i vores eksempel at blive ansat - ikke er påvirket af en beskyttet egenskab - som køn.

MÆND	Ukvalificeret	Kvalificeret
Afvis	15 (TN)	5 (FN)
Ansæt	20 (FP)	60 (TP)

KVINDER	Ukvalificeret	Kvalificeret
Afvis	60 (TN)	20 (FN)
Ansæt	5 (FP)	15 (TP)

Hvis 80 ud af 100 mandlige ansøgere og kun 20 ud af 100 kvindelige ansøgere bliver ansat i rekrutteringsscenariet, er den demografiske paritet ikke opfyldt. Det kan dog være acceptabelt, hvis de mandlige ansøgere faktisk er mere kvalificerede end de kvindelige ansøgere, som det ser ud til i dette tilfælde.



Idet der opnås retfærdighed på gruppeniveau, kan der opstå uretfærdighed på individniveau, da systemet kan fravælge ellers kvalificerede kandidater bare for at opnå demografisk paritet.

Projekt | AI i rekruttering

Forudsigelig paritet

I modsætning til demografisk paritet tager forudsigende paritet både klassifikatorens afgørelser og de sande resultater i betragtning. Den vurderer, om sandsynligheden for, at en ansøger er egnet til en stilling, er konsistent på tværs af forskellige grupper, forudsat at AI'en har udvalgt dem til ansættelse. Forventningen er, at denne sandsynlighed skal være næsten identisk for alle de grupper, der betragtes

MÆND	Ukvalificeret	Kvalificeret
Afvis	15	5
Ansæt	20	60

KVINDER	Ukvalificeret	Kvalificeret
Afvis	60	20
Ansæt	5	15

I rekrutteringseksemplet opfylder systemet kriterierne for forudsigende paritet, da tre fjerdedele af de ansatte ansøgere fra begge grupper faktisk er kvalificerede

Projekt | AI i rekruttering

Udlignede odds; falsk positiv rate & falsk negativ rate

Vedr. målingen, der er kendt som den **falske positive rate**, hævdes det, at sandsynligheden for, at en ukvalificeret kandidat fejlagtigt bliver ansat, bør være konsekvent på tværs af alle demografiske forhold.

MÆND	Ukvalificeret	Kvalificeret
Afvis	15	5
Ansæt	20	60

KVINDER	Ukvalificeret	Kvalificeret
Afvis	60	20
Ansæt	5	15

Vedr. målingen, der er kendt som den **falske negative rate**, hævdes det, at sandsynligheden for, at en kvalificeret kandidat fejlagtigt bliver afvist, bør være ensartet på tværs af alle demografiske forhold.

MÆND	Ukvalificeret	Kvalificeret
Afvis	15	5
Ansæt	20	60

KVINDER	Ukvalificeret	Kvalificeret
Afvis	60	20
Ansæt	5	15

Udlignede odds kombinerer disse to egenskaber ved at opfylde både falsk negativ rate-balance og falsk positiv rate-balance.

I rekrutteringseksemplet **opfylder** løsningen **ikke kriterierne**, da der er større sandsynlighed for, at de ukvalificerede mandlige ansøgere bliver ansat end de ukvalificerede kvindelige ansøgere, OG der er større sandsynlighed for, at de kvalificerede kvindelige ansøgere bliver afvist end de mandlige ansøgere.

Overvejelser om fairness-metrikker

- Et system kan ikke opnå forudsigende paritet, balance i falsk positiv rate og balance i falsk negativ rate på én gang på grund af matematiske begrænsninger.
- Afvejninger mellem forskellige retfærdighedsmål er iboende og skal håndteres med omtanke.
- Overvejelser om retfærdighed skal ses i konteksten af den specifikke applikation, de anvendes på.
- Ikke alle aspekter af retfærdighed kan kvantificeres, hvilket er viser retfærdighedens sociotekniske kompleksitet.

Projekt | AI i rekruttering

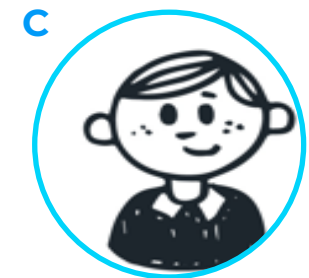
Individuel retfærdighed



Bachelorgrad
1 års relateret erhvervserfaring



Kandidatgrad
1 års relateret erhvervserfaring



Kandidatgrad
Ingen relateret erhvervserfaring

- Individuel retfærdighed adskiller sig fra gruppebaserede retfærdighedskriterier (de tidligere omtalte metrikker) ved at fokusere på behandlingen af enkeltpersoner.
- I dette koncept hævdes det, at personer med lignende karakteristika bør få sammenlignelige resultater.
- Det er komplekst at finde en metode til at måle individuelle ligheder.
- Forestil dig tre jobkandidater: Er A bedre i stand til at udføre jobbet end C?

Projekt | AI i rekruttering



IMPLEMENTERING AF MODELLEN



Informér ansøgere om brug af AI i rekrutteringen

Efter omhyggelige interne evalueringer er virksomheden klar til at lancere sit screeningsværktøj på nogle bestemte stillingsopslag for at indsamle den første feedback fra brugere og ansøgere.

Det er afgørende, at ansøgerne hele tiden er klar over, at de er i kontakt med et AI-drevet rekrutteringssystem, så tilliden dermed fremmes. Der bør skabes opmærksomhed og bevidsthed om fordelene ved AI-værktøjer helt fra starten, hvilket øger ansøgernes villighed til at bruge interaktiv teknologi som chatbots. Derudover er det vigtigt kontinuerligt at forbedre AI-rekrutteringsprocesserne og sikre klarhed om systemets funktion, da dette er afgørende for, at det bliver taget i brug og bliver en god hjælp for ansøgerne.

Det skal også sikres, at der er et menneske, der fører tilsyn og griber ind, når det er nødvendigt

Projekt | AI i rekruttering

Information til ansøgere om brug af AI i rekruttering | Eksempler på værktøjer

UNESCO's Ethical Impact Assessment

Et eksempel fra UNESCO's værktøj, Ethical Impact Assessment, der blev præsenteret tidligere

10.2.1. System Awareness:

- 10.2.1.1. Are users made fully aware when they are interacting with an AI system, as opposed to a human being?
- 10.2.1.2. Are individuals (directly or indirectly) impacted by the AI system made fully aware of when a decision (that impacted them) was informed by or made on the basis of an AI system or AI algorithms?
 - 10.2.1.2.1. Are they made aware of the extent to which they are impacted, including the rationale, benefits and limitations of the decision(s)?
- 10.2.1.3. Have appropriate explanations been put in place to help users and other impacted individuals understand the decision-making process or how the system works when required?
- 10.2.1.4. Have appropriate explanations been put in place to help the government bodies in charge of regulation understand the decision-making process or how the system works when required?
- 10.2.1.5. Has the decision to adopt the AI system been documented and communicated online?
- 10.2.1.6. Can the AI system make any decisions which the physical persons or legal entities in charge of the system lack expertise or competence to critique, modify or override?

Eccola Cards: en metode til at implementere etisk tilpassede AI-systemer

Data

#7 Privacy and Data

Motivation: Privacy is a rising trend in the wake of various recent data misuse reveals. People are now increasingly conscious about handing out personal data. Similarly, regulations such as the GDPR now affect data collection.

What to Do: Ask yourself:

- What data are used by the system?
- Does the system use or collect personal data? Why? How is the personal data used?
- Do you clearly inform your (end-)users about any personal data collection? E.g., ask for consent, provide an opportunity to revoke it etc.
- Have you taken measures to enhance (end-user) privacy, such as encryption or anonymization?
- Who makes the decisions regarding data use and collection? Do you have organizational policies for it?

Practical Example: Rather than collecting and selling data, appealing to privacy can also be profitable. Regulations are making it increasingly difficult to collect lots of personal data for profit. Privacy can be an alternate selling point in today's climate.

ECCOLA

©10-4601-20200415

Unesco, 2023; Vakkuri et al. 2021

Projekt | AI i rekruttering



DRIFT OG OVERVÅGNING



Efter udrulning

Mens interne vurderinger af retfærdighed blev gennemført før lanceringen, skal virksomheden også være klar til uafhængige eksterne revisioner.

Det er også en fordel løbende at efterspørge feedback efter implementeringen og etablere kanaler, hvor igennem man kan holde kontakten til interessenter. Dette indebærer at holde medarbejderne informeret og involveret i høringer og deltagelsesprocesser under hele implementeringen af AI-systemet i organisationen.

TAK



**Medfinansieret af
Den Europæiske Union**

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 EACEA kan holdes ansvarlig herfor.

2022-1-ES01-KA220-HED-000085257