

Ledelsesetiske udfordringer i forbindelse med indførelse af AI

Af Michael Jensen, direktør, Colea ApS
mj@colea.dk

Indhold

1. Introduktion
2. Udfordringerne kort fortalt
3. Fordomme forstærkes i AI
4. Fralæggelse af ansvar
5. Manglende transparens
6. AI kan gøre nogle jobs overflødige
7. Nogle gange er kolde beregninger ikke nok
8. Parkering af autonomi
9. Tillid
10. AI etisk tjekliste
11. Afrunding
12. Mere viden

1. Introduktion

Det tog Netflix 3,5 år at få 1 mio. brugere. Det tog Chat GPT 5 dage!

Med tsunamiagtig hastighed buldrer AI frem i verden i almindelighed og i organisationsverdenen i særdeleshed. Når noget udvikler sig så stærkt, er det ofte en god idé at træde et skridt tilbage og overveje fordele og ulemper – og ikke mindst sikre, at det vi gør, er etisk forsvarligt.

Hovedbudskabet i denne artikel er, at AI selvfølgelig skal bruges, men at det skal bruges med fornuft og omtanke. Overholder vi dette, vil vi kunne nå nye højder for vores organisation. Gør vi det *ikke*, risikerer vi at blive revet omkuld af tsunamien.

Som Cand. Mag i filosofi og med en 25-årig baggrund som virksomhedskonsulent, mener jeg at have en rimelig god basis for at kunne italesætte dette dilemma og formålet er IKKE at komme med en løftet pegefinger, men tværtimod at give nogle råd til, hvordan vi navigerer sikkert igennem de etiske udfordringer, der givet er i forbindelse med indførelse af AI.

Som læser, kan du forvente at få følgende ud af at læse denne artikel:

1. Kendskab til de væsentligste ledelsesetiske udfordringer
2. Inspiration og viden, der kan hjælpe med at navigere uden om eller igennem dem
3. I forlængelse af punkt 1 og 2 en større sandsynlighed for at opnå optimalt udbytte af indførelsen og brugen af AI

I næste kapitel vil jeg ganske kort komme ind på de væsentligste etiske udfordringer – for overblikkets skyld.

I de efterfølgende kapitler vil jeg gå lidt mere i dybden med udfordringerne for efterfølgende at sige noget om, hvordan vi håndterer dem. Denne opbygning burde gøre, at du kan nøjes med at slå ned på de udfordringer du selv imødeser – eller måske gøre det, når de uventet dukker op. Du er naturligvis også velkommen til at læse hele artiklen.

Jeg har fravalgt at tage emnet om datasikkerhed op – ikke fordi det er uvæsentligt eller ikke har en etisk dimension, men fordi jeg ved, der er fokus på denne del i EU's AI forordning. Jeg ønsker her at fokusere på de etiske udfordringer, der har en mellemmenneskelig etisk vinkel.

I den forbindelse kan det i øvrigt være vigtigt at pege på adskillelsen af etikens og juraens domæne: Nogle ting vi finder uetiske, som f.eks. at bagtale andre, er ikke ulovlige, og nogle ting, der er ulovlige f.eks. at køre for stærkt (et sted, hvor det ikke betyder en risiko for andre), finder vi (eller nogle af os) måske ikke uetiske.

En sidste kommentar inden jeg starter på selve artiklen er, at jeg har valgt at anvende termen "AI" bredt i stedet for at skulle skifte mellem forskellige termer – jeg håber på læserens tilgivelse for dette

2. Udfordringerne kort fortalt

Lad mig begynde med at opremse de ledelsesetiske udfordringer, jeg umiddelbart kan identificere. Der vil sandsynligvis dukke flere op efterhånden som AI breder sig, og der opstår nye anvendelsesmuligheder.

1. **Fordomme kan blive forstærket i AI:** Mennesker har fordomme og det er mennesker, der har udviklet og fodret AI med viden. Derfor er AI også biased og denne "biasedhed" kan blive forstærket via AI
2. **Fralæggelse af ansvar:** Når vi lægger beslutninger over på AI, er der en risiko - især når disse får negative konsekvenser - for at vi fralægger os selv ansvaret
3. **Manglende transparens:** Det kan være svært at gennemskue beslutningsprocesser truffet på baggrund af AI
4. **AI kan gøre visse jobs overflødige:** Dette kan skabe et ledelsesmæssigt dilemma

5. **Nogle gange er kolde beregninger ikke nok:** AI besidder ikke empati – et væsentligt karaktertræk hos gode ledere
6. **Parkering af autonomi:** Hvis vi i for høj grad lader AI overtage, risikerer vi at miste vores selvbestemmelse og muligvis også nogle af vores færdigheder
7. **Tillid:** Tillid er et nøglebegreb i al interaktion, denne kan blive udfordret af AI. En anden vinkel er, at vi mennesker har tendens til at have større tillid til maskiner end til mennesker, men AI begår nogle gange banale fejl, og kan endda narres til at rådgive om kriminalitet.

Dette er udfordringerne, som jeg ser dem nu. Som det givetvis er fremgået, er der visse steder overlap, men det må nødvendigvis være sådan.

I de efterfølgende kapitler kigger vi på udfordringerne en for en.

3. Fordomme forstærkes i AI

En af de ting, der kendetegner homo sapiens er, at vi hele tiden vurderer og fortolker det, vi perciperer. Hermed opbygges en række fordomme om verden og de ting, vi møder.

Dette bygger på en evolutionær trang til overlevelse og er dybest set en kombination af selvopholdelse, erfaring og frygt for det ukendte.

Udfordringen i forhold til AI er, at det er mennesker, der har fodret AI med den information, den så glimrende bruger – og denne information er i større eller mindre grad påvirket af vores fordomme.

Nogle af de fordomme vi kan se AI har, handler f.eks. om kønsroller og til dels også race. Ligeledes mener nogle, at Chat GPT politisk hælder til venstre for midten.

En helt generel udfordring i denne forbindelse er, at AI bygger på induktiv logik og maskinlæring. Dvs. hver gang tilfælde x opstår, øger det sandsynligheden for, at AI vælger x i sin rådgivning, hvilket igen øger sandsynligheden for x osv. Når samtidig AI ikke besidder emotioner/følelser, menneskelige egenskaber vi i dag ved fungerer som "intuitive" etiske indikatorer, er det klart, at AI kan risikere at overse etiske vinkler i sagen og alene træffe valg og rådgive ud fra objektiv sandsynlighed.

Den etiske udfordring består dels i risikoen for at fordomme, utilsigtet, kan blive forstærket ved brug af AI, dels i at vores beslutninger påvirkes af denne eller hin fordom, og at vi ubevidst overtager disse.

Hvordan kan vi så minimere sandsynligheden for at trække AIs fordomme med over i vores beslutningsprocesser?

Dette kan vi gøre via følgende:

1. Vær lige så kritisk over for AIs anbefalinger, som du ville være, hvis anbefalingen kom fra et menneske
2. Hav et klart defineret værdisæt som grundlæg for beslutninger generelt
3. Sørg for at udfordre AI i selve arbejdsprocessen, så AI selv kan medvirke til at nuancere sine forslag.

4. Fralæggelse af ansvar

Det kan være fristende at overdrage en del af det personlige ansvar for beslutninger til AI – specielt, når beslutningerne har mindre heldige konsekvenser.

Men uanset konsekvenserne må vi holde fast i, at det er os selv, der sidder med den endelige beslutning, og vi er derfor nødt til at beregne de mulige konsekvenser af de valg, vi træffer.

Inden for etikken arbejder vi med tre overordnede typer af konsekvenser, der kan være relevante at kigge på i denne forbindelse:

1. Forventede eller tilsigtede konsekvenser
2. Utilsigtede konsekvenser, som vi *ikke* kunne forudse
3. Utilsigtede konsekvenser, som vi *burde* kunne forudse

For kategori 1) har vi et naturligt etisk ansvar, mens det straks er sværere for kategori 2). Her vil vurderingen af en række subjektive forhold spille ind i vurderingen etisk ansvar eller ej. Denne akademiske diskussion ligger imidlertid uden for denne artikels formål, hvorfor jeg vil spare læseren for den.

Den kritiske kategori her er naturligvis 3) for her kan det hævdes, at vi - hvis vi havde tænkt os bedre om - burde *have* forudset at f.eks. noget kunne gå galt.

Derfor er det vigtigt i forbindelse med *alle* de beslutninger, vi træffer, at vi overvejer de tre typer af konsekvenser og også er klar til at påtage os ansvaret for dem. Det er i hvert fald ikke ansvarsfraskrivende at bygge sine beslutninger på AI.

Vi bør derfor *i alle beslutningsprocesser*, om vi konsulterer AI eller ej, påtage os det overordnede ansvar og være tydelige i vores kommunikation om dette.

5. Manglende transparens

Når vi kaster et spørgsmål op i f.eks. Chat GPT, kan det være svært at gennemskue hvilke præmisses, der ligger til grund for de svar, vi får, og som altid med manglende transparens kan motiverne bag givne beslutninger problematiseres.

Derfor må vi fokusere på åbenhed, når vi kommunikerer AI-baserede beslutninger og valg. Ligeledes er det vigtigt, at vi, så vidt muligt, fodrer AI med al nødvendig information, når vi benytter os af den.

Derfor kan det også anbefales, at der på højeste niveau i organisationen og gerne helt oppe på bestyrelsesniveau, hvis organisationen har en sådan, defineres et sæt klare politikker for anvendelse af AI. Herunder bør det overvejes, om der er særlige områder, hvor beslutninger har en så alvorlig og ansvarspådragende karakter, at det helt bør undgås at konsultere AI.

I denne forbindelse er det som med alle andre områder vi kommer ind på i artiklen, vigtigt at holde os for øje, at én ting er at udforme retningslinjer, politikker osv., noget helt andet er det at sikre, at disse også afspejles i *adfærden* i alle lag af organisationen.

6. AI kan gøre nogle jobs overflødige

At ny teknologi kommer til at koste job har været postuleret mange gange. Bl.a. da Internettet og digitale bøger fik skylden for at ødelægge bogmarkedet. Dette er ikke sket, hvilket imidlertid ikke et argument for, at det ikke kommer til at ske, at AI udkonkurrerer visse jobs.

Den israelske makrohistoriker Yuval Noah Harari argumenterer i sin bog *Homo Deus* for, at jo mere specialiseret et erhverv er, desto større er sandsynligheden for, at det kan blive udfordret eller udkonkurreret af AI.

Det kan f.eks. handle om chauffør erhvervet, advokatgerningen og visse dele af lægegerningen.

På denne måde at blive erstattet af en maskine er alt andet lige ikke rart, men jeg er overbevist om, og så må tiden dømmes om jeg har ret, at med det tempo nye jobs opstår i, så er der rigeligt nye muligheder på beskæftigelsesområdet.

Som jeg hørte det udtrykt for nylig

AI kommer ikke til at erstatte mennesker, men det kommer til at erstatte mennesker, der ikke anvender AI

Mikkel Næsager

Det er altså et åbent spørgsmål, om dette er et problem og, hvis, om det også er et etisk problem.

7. Nogle gange er kolde beregninger ikke nok

I forlængelse af det jeg nævnte ovenfor om AI's beregninger ud fra induktiv logik og sandsynlighed, kan det mellem menneskelige perspektiv risikere at blive nedtonet.

Et tænkt eksempel kan være den personalechef, der lader AI udvælge, hvem der står for i forbindelse med en fyringsrunde. Hvad enten vi gør dette uden forud bestemte udvælgelsesparametre eller med disse, så risikerer vi her, at et vigtigt ledelsesmæssigt perspektiv går tabt: Perspektivet på det enkelte individ af kød og blod.

Ledelse handler om mennesker

Et faktum vi aldrig må negligere. Derfor bør vi være meget påpasselige med hvilke elementer, vi ønsker at involvere AI i: Det kunne være fristende i forbindelse med en ubehagelig ledelsesbeslutning at konsultere AI, men vi skal passe på, at vi ikke her bevæger os ud på en etisk glidebane, hvor vi f.eks. i konfliktskyhed forsøger at undgå upopulære eller "svære" beslutninger ved at lade AI træffe dem.

Igen bør klare etiske politikker afgrænse det område, hvor ledelsen må konsultere AI.

Jeg vil anbefale at vi, som minimum, undgår at lade AI træffe beslutninger om

- a) Ansættelse
- b) Afskedigelse
- c) Hvem vi skal give en given opgave
- d) Personfølsom kommunikation

Det må være op til den enkelte organisation at etablere et sæt klare rammer for dette.

8. Parkering af autonomi

Det er helt afgørende for vores intuitive selvforståelse at opfatte os som og at kunne fungere som autonome væsener.

Autonom kommer af oldgræsk *auto*, "selv" og *nomos* "lov", dvs. "selvlov-givende væsener.

Den tyske filosof *Immanuel Kant* er den første til for alvor at definere rammen for det selvlovgivende væsen. Ifølge Kant forudsætter autonomi, at vi agerer ud fra et fast sæt af regler eller principper, og at vi respekterer vores medmennesker som havende deres egne mål, samt at vi aldrig betragter dem som blot og bare midler til opnåelse af vores egne mål.

Hvis vi imidlertid parkerer væsentlige beslutninger, og her er "væsentlige beslutninger", *enhver beslutning, der har etisk relevans for organisationer*

eller individer, hos AI, så mister vi vores autonomi i den udstrækning, vores principper og etiske fundament risikerer at blive udvandet.

Hvorfor det kunne man fristes til at spørge?

Fordi autonomibegrebet hænger uløseligt sammen med ideen om viljens frihed.

Fri vilje giver ingen mening, hvis ikke vi kan binde den op på en identitet, og identitet hænger sammen med en indre kerne bl.a. bestående af et mere eller mindre klart defineret sæt af værdier eller principper, der sætter en grænse mellem det, vi ønsker, og det, vi finder værdifuldt. I samme grad som denne grænse udviskes, mister vi vores autonomi.

Derfor er det centralt, at vi ikke må lade brugen af AI komme til at undergrave vores principper for god adfærd.

Den bedste måde at undgå dette på er at have et klart defineret værdisæt. Her mener jeg ikke det "pynteværdisæt" de fleste organisationer har hængende i kantine eller på deres hjemmeside, men noget meget skarpe og klart forankret, som er rettesnoren for vores valg.

9. Tillid

Mennesket er et socialt dyr, og for at vi skal kunne fungere sammen, er *tillid* en central ting. At jeg, med den danske filosof *K.E. Løgstrups* ord, kan have tillid til, at når jeg udleverer mig selv til dig, så tager du det op i samme tone, som jeg udleverer mig *uden at tage ansvaret for mig selv fra mig*.

Umiddelbart møder vi andre mennesker med tillid – indtil de forbryder sig imod den.

For at en organisation skal kunne fungere, er det selvsagt nødvendigt, at vi har en udstrakt grad af tillid til hinanden, og vi skal passe på med, at brugen af AI ikke nedbryder denne eller dele deraf.

Dette kan f.eks. ske, hvis en leders brug af AI beslutningsprocesser er ude af trit med den måde, lederen normalt agerer på. Omvendt kan det, som tidligere nævnt, være farligt at nære overdreven tillid til beslutninger truffet på baggrund af AI.

Der er altså grundlæggende to typer af tillidsmæssige udfordringer i forbindelse med brug af AI. Disse er vi nødt til at minimere risikoen for en realisering af.

Her kommer både den sunde dømmekraft og, som nævnt flere gange, et sæt af politikker eller rammer ind i billedet.

10. AI-etisk tjekliste Jeg har gennemgået mange ting ovenfor og som en hjælp til den organisation, der skal i gang med AI, kunne følgende AI-etiske tjekliste måske være en hjælp:

1. Forståelse af AI og dens begrænsninger:

- Er der en grundlæggende forståelse af, hvad AI kan og ikke kan?
- Er der klarhed over, hvor AI's beslutninger kommer fra og hvilke data, der fodrer disse beslutninger?

2. Identifikation af fordomme:

- Er der foretaget en gennemgang for at identificere potentielle fordomme i de data, der bruges til at træne AI?
- Er der implementeret procedurer for løbende at overvåge og rette fordomme i AI-systemer?

3. Ansvar og gennemsigtighed:

- Er der klare retningslinjer for, hvordan og hvornår AI skal bruges/ikke bruges i beslutningsprocesser?
- Er der etableret en proces for at sikre ansvarlighed og gennemsigtighed i beslutninger taget med hjælp fra AI?

4. Ethiske retningslinjer og politikker:

- Er der udviklet et sæt af etiske retningslinjer specifikt for brugen af AI – *i denne organisation*?
- Er disse retningslinjer kommunikeret klart til alle medarbejdere og relevante parter?

5. Uddannelse og bevidsthed:

- Er der tilbudt uddannelse og ressourcer til medarbejdere om etisk brug af AI?
- Er der skabt bevidsthed om de etiske udfordringer og dilemmaer forbundet med AI?

6. Overvågning og evaluering:

- Er der etableret procedurer for løbende overvågning af AI's performance og etiske konsekvenser?
- Er der planlagt regelmæssige evalueringer af AI-systemernes indvirkning på organisationen og dens stakeholders?

7. Håndtering af jobændringer:

- Er der strategier på plads for at håndtere de jobændringer, som AI kan medføre?
- Er der overvejet omskoling eller andre støtteforanstaltninger for medarbejdere, hvis roller ændres på grund af AI?

8. Tillid og relationer:

- Er der taget skridt til at opbygge og vedligeholde tillid blandt medarbejdere i forhold til brugen af AI?
- Er der overvejelser om, hvordan AI påvirker interpersonelle relationer og organisationens kultur?

Dette kunne være et bud på en tjekliste. Denne kan med fordel tilpasses den konkrete organisation.

11. Afrunding

Jeg har i denne artikel forsøgt at udpege nogle af de væsentlige ledelsesetiske udfordringer i forbindelse med indførelse af AI.

Budskabet var fra starten at vi skal selvfølgelig bruge AI, men vi skal samtidig sikre os imod at falde i de etiske faldgruber, denne brug medfører.

Jeg har i forbindelse med beskrivelse af de forskellige etiske udfordringer udpeget forskellige ting, vi kan gøre for at imødegå dem.

Helt overordnet tegner der sig fire veje til at håndtere de etiske udfordringer:

1. Udpeg de mulige etiske udfordringer på forhånd. Bevidsthed er ofte den bedste vej til at undgå udfordringer+.
2. Udform en AI politik og en række AI-etiske grundprincipper
3. Sørg for løbende at lade den sunde dømmekraft få plads i beslutningsprocesserne
4. Sørg for at sikre at ovenstående transformeres i egentlig adfærd i organisationen – evt. via træning af nøglemedarbejdere.

12. Mere Viden

<https://videnskab.dk/teknologi/er-chatgpt-venstreorienteret-forskere-har-undersoegt-14-chatbotters-politiske-staasted/>

https://videnskab.dk/teknologi/glem-de-dystopiske-fremtidsscenerier-ai-er-allerede-udbredt-i-dag-men-risici-er-ofte-skjulte/?utm_medium=email&utm_source=vores-nyhedsbrev

K. E. Løgstrup (1956): *Den etiske fordring*, Gyldendals Forlag

Ledelsesetiske udfordringer i forbindelse med indførelse af AI

Youval Noah Harari (2016): *Homo Deus – En kort historie om i morgen*,
Lindhardt og Ringhoff