#9

MAJ 2026

MÅNEDENS SINGULARITET:

DARTHMOUTH

SOMMEREN HVOR INTELLIGENS BLEV ET PROJEKT

AIPortalen.

Det seriøse medie om AI

TEMA:

NÅR MASKINEN VÆLGER

AI, BIAS OG DEN NYE KAMP OM JOURNALISTIKKENS DØMMEKRAFT

AI PÅ MARIENBORG

DA KUNSTIG INTELLIGENS BLEV REGERINGSPOLITIK

PROJEKT Y VIL IKKE BARE LAVE ET MEDIE

DE VIL EJE MASKINEN BAG

Indhold

LEDER: BIAS ER IKKE FEJLEN. DET ER RETNINGEN

AI-debatten har længe behandlet bias som en teknisk fejl. Noget, der kan opdages, renses ud og rettes i næste version. Men måske er det den forkerte måde at forstå problemet på.

For bias er ikke kun det, der sker, når AI svarer stereotypt, udelader bestemte perspektiver eller gengiver verden skævt. Bias er også den vægtning, der gør et svar muligt. Hvad tæller som relevant? Hvad bliver fremhævet? Hvad forsvinder som støj? Hvilke kilder, værdier og antagelser får lov til at fremstå som neutrale?

Det er den tråd, der løber gennem AI Portalen #9.

Vi begynder i medierne, hvor AI ikke længere kun er et eksperiment i kanten af redaktionen. Kortlægningen viser, at danske medier allerede bruger AI til research, transskribering, rubrikker, versionering, personalisering, kvalitetsmåling og interne værktøjer. Men den viser også, at åbenheden halter efter praksis. AI er på vej ind som redaktionel infrastruktur — og dermed også som en ny kilde til usynlig vægtning.

Projekt Y gør spørgsmålet endnu tydeligere. Her er AI ikke bare et værktøj i journalistikken, men selve platformen bag et kommende medie. Y.dk er ansigtet udadtil. Y-Rails er maskinen bag. Og hvis journalistikken i stigende grad bliver én funktion i et større teknologisk system, bliver det afgørende spørgsmål ikke kun, hvad der publiceres, men hvem der ejer og kontrollerer maskinen bag.

Derfor handler dette nummer ikke kun om AI i medier. Det handler om dømmekraft.

Peder Hammerskov fra DMJX formulerer udfordringen præcist: Den svære del er ikke længere at få journalister til at bruge AI. Det er at få dem til at bruge den reflekteret. “Human in the loop” lyder betryggende, men betyder kun noget, hvis mennesket faktisk har tid, viden og kritisk sans nok til at gribe ind.

Det er derfor, bias ikke kan reduceres til en teknisk underkategori. Bias er dér, hvor magten viser sig: i data, modeller, institutioner, redaktionelle valg, politiske problemforståelser og teknologiske infrastrukturer.

AI kan hjælpe os med at se verden klarere. Men kun hvis vi holder op med at lade, som om maskinen ser fra ingen steder.

Bias er ikke altid fejlen i systemet.

Ofte er bias selve retningen.

Illustration: Genereret med ChatGPT

TEMA: NÅR MASKINEN VÆLGER

KORTLÆGNING: SÅDAN BRUGER DANSKE MEDIER AI LIGE NU

Sådan bruger danske medier AI lige nu

Af Mark Sinclair Fleeton

AI er ikke længere et sideprojekt i danske medier. Det bruges allerede til research, transskribering, rubrikker, versionering, personalisering, kvalitetsmåling og interne værktøjer. Men åbenheden om, hvor langt medierne er gået, halter stadig bagefter.

Hvis man kun lyttede til den offentlige debat, kunne man tro, at danske medier stadig står og tester AI forsigtigt ude i periferien. Det gør de ikke. AI er allerede rykket ind i hverdagen. Ikke som førerløse redaktioner eller helautomatiske medier, men som noget mere lavpraktisk og derfor måske også mere betydningsfuldt: research, transskribering, rubrikker, resuméer, dataudtræk, SoMe-tekster, nyhedsbrevstekst og andre former for produktionsstøtte.

Det mere præcise billede er derfor ikke, at danske medier enten bruger eller ikke bruger AI. Det mere rigtige er, at de befinder sig på forskellige trin i samme bevægelse: fra eksperimenterende hverdagsbrug til interne toolkits, personalisering, kvalitetsmåling og mere systematisk workflowstyring.

Den foreløbige kortlægning bygger på offentligt tilgængelige cases, præsentationer, retningslinjer og rapporter. Den er ikke en fuldstændig opgørelse over al AI-brug i danske medier, men viser de mest synlige og veldokumenterede mønstre. Samlet tegner materialet et billede af en branche, hvor AI allerede er blevet en del af den daglige drift flere steder, men hvor den offentlige gennemsigtighed stadig er ujævn.

AI er blevet hverdag

Det stærkeste fælles udgangspunkt kommer fra DMJX's forsknings- og udviklingsarbejde og fra rapporten *AI og trivsel blandt danske journalister* fra Teknologisk Institut og Velliv Foreningen. Begge peger i samme retning: Implementeringen varierer meget fra mediehus til mediehus, men AI er allerede blevet en del af de daglige arbejdsgange flere steder. Rapporten fremhæver især transskribering, research, rubrikgenerering og korrektur som udbredte anvendelser.

JP/Politikens Hus: AI som system, ikke bare værktøj

Skal man pege på en dansk frontløber, er JP/Politikens Hus en af de tydeligste. Offentlige beskrivelser af MAGNA, PIN-projektet og JP/Politikens fælles AI-enhed tegner et billede af et mediehus, der ikke nøjes med enkeltstående AI-værktøjer, men er i gang med at bygge egentlig AI-infrastruktur omkring redaktionen.

JP/Politikens fælles AI-enhed, JPPOL AI, blev dannet i begyndelsen af 2024 og bygger videre på erfaringer fra PIN-projektet. Enheden består af 11 fuldtidsmedarbejdere og fire deltidsansatte og rummer blandt andet ML-specialister, engineers, data scientist og UI/UX-kompetencer. Det siger noget om ambitionsniveauet: AI er ikke kun placeret som et redaktionelt eksperiment, men som et produkt- og teknologiområde i koncernen.

Arbejdet er organiseret omkring tre primære produktområder: generativ AI, personalisering/ recommendation systems og metadata. Generativ AI handler om redaktionelle værktøjer, der kan hjælpe med rutineopgaver og understøtte rigere nyhedsdækning i tekst, billede og lyd. Recommendation-sporet handler om personalisering af nyhedsoplevelsen, redaktionelt kontrolleret og baseret på indsigter i personaliserede nyhedsstrømme. Metadata-sporet handler om at skabe rigere metadata om indhold i tekst, billede og lyd, så indhold kan aktiveres på nye måder og give dybere indsigt i, hvordan det bliver brugt.

MAGNA er et centralt eksempel på den udvikling. Ifølge materialet blev EB MAGNA udviklet i andet halvår 2023 og omfatter blandt andet artikelværktøjer, Live Center, RAG-løsningen "Spørg Arkivet" og feedbackfunktioner. I praksis betyder det, at journalister får adgang til en chatrobot i et sikkert miljø, semantisk søgning på tværs af koncernens artikelarkiver, mulighed for at generere indhold på baggrund af nyhedsmediernes egne arkiver og artikelværktøjer, der kan hjælpe med rutineopgaver og artikelskrivning.

Eksempler på MAGNAs anvendelsesområder omfatter brug af arkivsøgning som input til faktabokse, MAGNAs bud på faktabokse, rubrikforslag, kritisk gennemlæsning, spørgsmål til ubesvarede vinkler og omskrivning til artikeludkast. Det er dermed ikke kun et chatværktøj, men en samling redaktionelle funktioner, der kobler generativ AI til koncernens egne arkiver og arbejdsprocesser.

Men implementering af AI-systemer handler ikke kun om teknologi. JP/Politikens arbejder med feedback fra brugerne, testbrugere på redaktionerne, workshops, redaktionsbesøg og undervisning. Ifølge JPPOL AI har omkring halvdelen af feedbacken været positiv, og erfaringerne peger blandt andet på tre centrale udfordringer: Generativ AI til nyheder skal baseres på eksterne fakta, gennemsnitligt sprogbrug er en hovedudfordring, når tekster skal passe til en redaktionel profil, og brugervenlighed samt integration i eksisterende arbejdsprocesser er afgørende for bred adoption.

MAGNAs næste generationer peger på, hvor udviklingen kan bevæge sig hen: versionering af nyheder i tekst, lydversioner af nyhedshistorier, chat med artiklen, temaet eller arkivet, illustrationer — dog med fotorealisme som "no go" — og research. Samlet gør det JP/Politikens Hus til mere end en case om AI i redaktionen. Det ligner et mediehus, hvor AI er på vej ind som en del af både produktion, distribution, metadata, personalisering og brugeroplevelse. AI er ikke kun et værktøj ved skrivebordet. Det er ved at blive en del af infrastrukturen.

Sjællandske Medier: når AI bliver en arbejdsvane
Hvor JP/Politikens Hus viser den tunge systemintegration, viser Sjællandske Medier noget andet: hvordan AI kan blive en driftsteknologi. Ved en konference på DMJX i januar 2025 præsenterede Chefredaktør Carsten Lysdal værktøjet Chatty som hjælp til transskribering, versionering og SoMe-tekster. Lysdal beskrev AI som noget, der har frigivet mere tid til journalistikken.

Det gør Sjællandske Medier til et vigtigt eksempel. Her er AI ikke først og fremmest et stort strategisk narrativ, men noget, der flytter rundt på de daglige arbejdsgange. Det viser, hvordan AI i medier hurtigt bliver mest afgørende, når den holder op med at være spektakulær og i stedet bliver rutine.

TV 2 og regionerne: den tydeligste governance
Hvis JP/Politikens Hus og Sjællandske Medier især viser, hvordan AI bruges, viser TV 2 Danmark og TV 2-regionerne noget andet: hvordan den styres. De fælles retningslinjer for brug af generativ AI i nyhedsjournalistik og dokumentarer er blandt de tydeligste offentlige dokumenter på området i Danmark. Her fremgår det blandt andet, at AI-genereret eller markant AI-ændret indhold skal mærkes tydeligt, at der skal være menneskelig kontrol, og at man som udgangspunkt ikke udgiver fotorealistisk visuelt materiale skabt med generativ AI i nyhedsjournalistik. Hele artikler udgives ikke uden journalistisk bearbejdning, og eksperimenter med chatbots og opsummeringer uden fuld redaktionel godkendelse skal som minimum deklareres.

TV 2 er ikke nødvendigvis længst fremme på det teknisk område. Men de kan fremvise den mest synlige danske model for governance, mærkning og ansvar. Det er i sig selv væsentligt.

Altinget: mindre hype, mere data
Altinget peger i en lidt anden retning. I et nyere LinkedIn-opslag skriver ansvarshavende chefredaktør Jakob Nielsen, at Altinget bruger AI "*mange steder*", men at den "*virkelige satsning*" ligger inden for datajournalistik, blandt andet i arbejdet med meningsmålinger og vælgervandringer.

Det peger på en mere lavmælt strategi. Ikke AI som brandinghistorie, men AI som støtte til analyse, strukturering og datatung journalistik. Det er en vigtig nuance, fordi den viser, at AI i medier ikke nødvendigvis kommer med fanfare. Nogle steder glider den bare ind i de dele af journalistikken, hvor store datamængder og mønstre i forvejen fylder meget.

Kvalitetsmåling: når AI også vurderer journalistikken
En mindre synlig, men principielt vigtig brug af AI handler om kvalitetsmåling. Her bruges teknologien ikke først og fremmest til at producere mere indhold, men til at analysere det indhold, medierne allerede laver.

Constructive Institute har udviklet værktøjer, der bruger AI og maskinlæring til at analysere journalistik ud fra konstruktive kriterier. Constructive News Mirror beskrives som en platform, der måler journalistisk kvalitet og giver redaktioner ugentlige rapporter om, hvor konstruktiv deres dækning har været.

Constructive News Algorithm beskrives som et maskinlæringsværktøj, der bruger sproganalyse til at identificere konstruktive elementer og bias i nyhedsdækning. Værktøjet er udviklet som et supplement til den redaktionelle proces og skal hjælpe redaktioner med at evaluere, hvor konstruktiv deres dækning er.

AI bruges ikke kun som skrivehjælp, researchassistent eller produktionsstøtte. Den kan også blive et redaktionelt spejl, der måler mønstre i dækningen og gør det lettere at diskutere, om journalistikken faktisk lever op til de værdier og kvalitetsmål, redaktionen selv har sat.

Mere end effektivisering
Samlet viser kortlægningen, at AI i medier ikke længere kun handler om at gøre gamle opgaver hurtigere. Brugen spænder fra produktionsstøtte og research til personalisering, versionering, styring, dataanalyse og kvalitetsmåling.

Det udvider billedet af AI i journalistikken. Teknologien er ikke kun på vej ind som medhjælper ved skrivebordet. Den er også på vej ind som system, analyseværktøj og styringsredskab — og i stigende grad som noget, der former både arbejdsgange, redaktionelle prioriteringer og mødet mellem medie og bruger.

Mangel på transparens
Måske er det vigtigste fund ikke, hvor meget medierne bruger AI, men hvor lidt de fortæller om det. Reuters Institute's Danmark-kapitel fra 2025 konstaterer, at danske nyhedsorganisationer investerer tungt i generative AI-værktøjer og gradvist implementerer dem i de daglige nyhedsoperationer. Men det betyder ikke, at offentligheden får meget at vide om, hvordan det sker.

Her er TV 2's retningslinjer undtagelsen snarere end reglen. For selv i de cases, hvor brugen er veldokumenteret, er det ofte uklart, præcis hvor langt medierne er gået, hvilke værktøjer der bruges hvor, og hvilke grænser de sætter internt. Det gælder især uden for de medier, der selv aktivt har valgt at fortælle om deres AI-praksis. Det gør transparens til en historie i sig selv. Ikke bare hvordan medier bruger AI, men hvor åbne de er om det.

Publikum er mere skeptisk end medierne
Der er også en anden grund til, at den åbenhed betyder noget: tillid. Reuters Institute's 2025-rapport peger på, at publikums skepsis over for AI i nyheder stadig er høj på tværs af lande, og Danmark ligner ikke et marked, hvor den modstand er på vej væk. Tidligere danske målinger har også vist betydelig utryghed ved journalistik, der i høj grad er produceret ved hjælp af AI.
Det gør AI i medier til mere end et internt produktionsspørgsmål. Det er også et tillids- og transparensspørgsmål. Hvis medierne vil bruge AI dybere i deres arbejde, kommer de også til at skulle forklare bedre, hvor, hvordan og hvorfor de gør det.

Fra værktøjer til infrastruktur
Den foreløbige kortlægning viser, at danske medier især bruger AI som produktionsstøtte, researchværktøj, dataanalyse, kvalitetsmåling, personalisering og interne redaktionelle systemer. Men udviklingen stopper næppe dér.

Et projekt som Projekt Y peger på næste trin i samme bevægelse: et medie, der ikke bare bruger AI i enkelte led af arbejdet, men bygger hele produktet oven på en platform, hvor overvågning, workflow, generering, distribution og på sigt licensering hænger sammen.

Projekt Y er ikke i drift endnu, og modellen er ikke bevist. Men som case er projektet interessant, fordi det viser, hvordan AI i medier kan bevæge sig fra

værktøj til infrastruktur. Hvor de etablerede medier i høj grad forsøger at lægge AI oven på eksisterende organisationer, forsøger Projekt Y at bygge organisationen omkring AI fra begyndelsen.

En branche på forskellige trin

Den foreløbige kortlægning peger på, at danske medier ikke længere står uden for AI-udviklingen. De er allerede i gang med at opbygge et nyt mellemstadie, hvor AI bliver en usynlig eller halvsynlig medproducent i dele af arbejdsprocessen.

Forskellen mellem medierne ligger især i fire ting: hvor integreret teknologien er, om AI primært bruges til produktion, analyse eller infrastruktur, hvor tydelige grænser de sætter, og hvor åbne de er over for publikum om brugen.

Det mest præcise billede er derfor ikke, at danske medier enten bruger eller ikke bruger AI. Det mere rigtige er, at de befinder sig på forskellige trin i samme bevægelse — fra eksperimenterende hverdagsbrug til interne toolkits, personalisering, kvalitetsmåling og begyndende AI-infrastruktur. Og måske er det netop dér, den vigtigste historie ligger: ikke i om AI kommer ind i danske medier, men i at den allerede er der. Spørgsmålet er ikke længere kun, hvad den kan. Spørgsmålet er, hvor meget af den der foregår uden for offentlighedens blik.

FAKTABOKS

SÅDAN ER KORTLÆGNINGEN LAVET

Denne artikel bygger på en journalistisk kortlægning af offentligt tilgængelige oplysninger om brugen af AI i danske medier.

KORTLÆGNINGEN OMFATTER:

- Offentlige rapporter og analyser om AI i mediebranchen
- Præsentationer og oplæg fra DMJX's AI-konference i januar 2025
- Offentlige beskrivelser af konkrete mediecases
- Publicerede retningslinjer for brug af AI
- Offentlige udtalelser fra medieledere og redaktionelle profiler

MATERIALET BYGGER BLANDT ANDET PÅ:

- **JP/Politikens Hus** / MAGNA / PIN-projektet
- **Sjællandske Medier** / Chatty
- **TV 2 Danmark og TV 2-regionernes** AI-retningslinjer
- **Altingets** offentlige udmeldinger om AI og datajournalistik
- **DMJX's** forsknings- og udviklingsarbejde om AI i journalistik
- Rapporten ***AI og trivsel blandt danske journalister*** fra Teknologisk Institut og Velliv Foreningen

HVAD KORTLÆGNINGEN KAN SIGE

Kortlægningen kan bruges til at vise:

- hvilke former for AI-brug der er offentligt dokumenteret
- hvilke mediehuse der fremstår som frontløbere eller tydelige cases
- hvilke mønstre der tegner sig på tværs af branchen
- hvor der findes offentlige retningslinjer, værktøjer og strategier

HVAD KORTLÆGNINGEN IKKE KAN SIGE SIKKERT

Kortlægningen er ikke en fuldstændig opgørelse over al AI-brug i danske medier. Den siger derfor ikke præcist:

- hvor meget AI alle danske medier bruger internt
- hvilke værktøjer der anvendes i alle redaktioner
- hvor langt hvert enkelt medie er i praksis
- hvordan alle interne workflow, kontrolmekanismer og begrænsninger ser ud

FORBEHOLD

Kortlægningen er stærkest dér, hvor medier selv har offentliggjort beskrivelser, retningslinjer eller præsentationer. Det betyder også, at medier med høj offentlig transparens lettere fremstår tydelige i materialet end medier, der arbejder mere lukket. Artiklen beskriver derfor først og fremmest den dokumenterede og synlige del af udviklingen.

PEDER HAMMERSKOV: DEN SVÆRE DEL ER IKKE AT FÅ JOURNALISTER TIL AT BRUGE AI - MEN AT FÅ DEM TIL AT BRUGE DET REFELEKTERET

AF MARK SINCLAIR FLEETON

AI er rykket ind i journalistikken. Ikke som den førerløse redaktion, hvor mennesker er skrevet ud af ligningen, men som en voksende del af det daglige arbejde. Research, transskribering, idéudvikling, rubrikker, korrektur, opsummeringer og interne redaktionelle værktøjer. Men for Peder Hammerskov, leder af Center for AI på DMJX - Danmarks Medie- og Journalisthøjskole, er det ikke længere nok at konstatere, at AI bliver brugt. Det afgørende spørgsmål er, hvordan den bliver brugt — og om journalisterne forstår, hvad de gør, når de bruger den. Det svære er at skabe en kultur, hvor AI bruges ansvarligt, kritisk og med journalistisk dømmekraft.

"*Der er mange flere, der er begyndt at bruge AI til alle mulige ting i deres journalistik. Mange flere har fået øjnene op for, hvad det kan. Men i dansk kontekst er alle medierne stadig i gang med at finde ud af, hvordan de gør det her, og hvad de egentlig skal med det,*" siger han.

Han beskriver en branche, der både er længere fremme, end mange måske tror, og samtidig mindre forandret, end mange forestillede sig, da generativ AI for alvor slog igennem.

"*Man kan godt synes, det går langsomt, men AI fylder jo alle steder. Det er den største snak inden for innovation, medieinnovation og medieuddannelse,*" siger han.

Fra tidsbesparelse til faglig sparring
Den første bølge af AI i medierne handlede i høj grad om effektivisering. Det var oplagte opgaver, der blev automatiseret eller understøttet: transskribering, korrektur, resuméer, rubrikforslag og hjælp til at bearbejde store tekstmængder. Det spor findes stadig. Men ifølge Peder Hammerskov er der også sket en forskydning.

"*Jeg tror, vi er kommet længere ud af det rene effektiviseringsparadigme og over i noget, hvor AI mere bliver en kvalitetssparringspartner,*" siger han.

Det er en vigtig ændring. Hvis AI kun bruges til at spare tid, kan man diskutere teknologien som et produktionsværktøj. Men hvis AI begynder at fungere som sparringspartner, bevæger den sig tættere på selve journalistikkens kerne: vinkling, research, prioritering, sprog, kildevalg og kritisk tænkning.

Det er her, Peder Hammerskov mener, at diskussionen bliver mere krævende. For teknologien kan hjælpe journalister med at tænke bredere, stille nye spørgsmål og opdage blinde vinkler. Men den kan også udviske nuancer, indføre fejl, forstærke bias og gøre journalisten mindre opmærksom på sine egne valg. Derfor er det afgørende, at AI ikke bare bliver noget, man "bruger", men noget man lærer at forholde sig til.

Det handler ikke kun om teknologi, men om kultur Peder Hammerskov ser flere danske medier bevæge sig fra løs brug af åbne værktøjer mod mere sikre og interne AI-systemer. De mest modne mediehuse forsøger ikke kun at stille værktøjer til rådighed, men at bygge AI ind i arbejdsgange, arkiver og redaktionelle systemer. Men han advarer mod at tro, at implementering først og fremmest handler om teknik.

Der er mange følelser forbundet med AI i journalistikken, understreger han. Frygt for ens egen jobsituation. Frygt for, om ens faglighed mister værdi. Usikkerhed om, hvad kollegerne tænker. Og bekymring for, hvad teknologien betyder for journalistikken og samfundet. Derfor handler implementering også om psykologisk tryghed.

"*Der er masser af følelser i det. Det gælder om at få skabt den tryghed, hvor man ved, at det er okay at bruge AI, hvor ledelsen har sat tydelige hjørneflag op for, hvad man må og ikke må, og hvor man ikke får hugget hovedet af, hvis man laver små fejl.*"

Det er en pointe, han også ser som central for mediernes AI-arbejde: Teknologien kan ikke bare lægges oven på redaktionerne. Den skal oversættes til kultur, praksis og fælles sprog.

Reglerne findes — men kender journalisterne dem?
Mange medier har efterhånden formuleret retningslinjer for brug af AI. Men ifølge Peder Hammerskov er der forskel på at have regler og på at have regler, der faktisk lever i organisationen.

"*Reglerne er der, men de kender dem ikke nødvendigvis. De er ikke altid gode nok til at formidle retningslinjerne, så de bliver en del af kulturen og den måde, man almindeligvis arbejder på,*" siger han.

Det skaber en gråzone. Nogle medier tillader, at AI bruges til artikeludkast. Andre tillader kun sparring, idéudvikling eller sproglig bearbejdning. Nogle steder må AI gerne hjælpe med struktur og overblik, men ikke producere brødtekst. Andre steder er rammerne mere åbne.

"*Der sker masser af skyggebrug derude. Folk bruger værktøjer, som de egentlig ikke må bruge, men som bare er det nemmeste, fordi de er vant til at bruge dem.*"

Det er ikke nødvendigvis udtryk for oprør eller ligegyldighed. Det kan også være et tegn på, at de ansvarlige systemer ikke er gode nok, eller at retningslinjerne ikke er blevet gjort relevante i hverdagen. For Peder Hammerskov viser det, at AI-governance ikke kun er et spørgsmål om dokumenter. Det er et spørgsmål om praksis.

Journalisterne er delte
Journalister reagerer meget forskelligt på AI. Nogle ser teknologien som en mulighed for at bruge mindre tid på rutineopgaver og mere tid på kilder, research og idéudvikling. Andre frygter, at AI gradvist flytter journalistisk kontrol fra mennesker til systemer.

"*Nogle er vildt bange for, hvad det kommer til at gøre ved håndværket. Kommer vi til at have journalistik, hvor mennesker afgiver for meget kontrol? Hvor sprogmodellerne langsomt beslutter, hvad vi skriver om, og hvordan vi skriver om det?*"

Samtidig møder han også journalister, der ikke nødvendigvis ser skriveprocessen som den vigtigste del af journalistikken.

"*Der er også nogle, der siger: Jeg synes egentlig ikke, det er særligt fedt at skrive. Jeg vil gerne ud og tale med mennesker og finde de vigtige historier. Hvis jeg kan slippe for noget af skriveprocessen, er det fint.*"

Det gør diskussionen vanskelig. AI rammer ikke kun produktionen, men også journalisters selvforståelse. For nogle er sproget og formen selve håndværket. For andre ligger journalistikkens værdi først og fremmest i researchen, kildearbejdet, dømmekraften og evnen til at finde historier.

Human in the loop er svært i praksis
De fleste medier vil fastholde, at mennesker stadig har ansvaret. AI kan hjælpe, men journalisten eller redaktøren skal kontrollere og godkende. Det lyder enkelt. Men ifølge Peder Hammerskov er det langt mere kompliceret i praksis.

"*Det er nemt at sige human in the loop, men svært at gøre i praksis. Hvordan sikrer vi, at vi reelt tager stilling til det, der kommer ud af systemet — også når vi har travlt?*" siger han.

Det er en af hans stærkeste advarsler. For menneskelig kontrol fungerer kun, hvis mennesket faktisk har tid, viden og opmærksomhed nok til at kontrollere. Hvis outputtet ser plausibelt ud, hvis deadline presser, og hvis værktøjet ofte leverer brugbare forslag, kan kontrollen hurtigt blive overfladisk.

Det er her, risikoen for automation bias opstår: at mennesker får for stor tillid til systemets forslag, fordi de ser professionelle, flydende og overbevisende ud. Peder Hammerskov er bekymret for, at begejstringen for teknologien nogle steder kan føre til for lidt kritisk distance.

"*Der er lidt for meget ureflekteret brug af det. Der er lidt for mange, der ikke helt forstår, hvor farligt det egentlig er — eller hvor dårligt det egentlig er.*"

Pointen er ikke, at AI ikke kan bruges. Pointen er, at brugen kræver mere faglig bevidsthed, ikke mindre.

Bias kan også blive en anledning til bedre journalistik
Bias er et af de problemer, Peder Hammerskov mener, medierne skal tage alvorligt. Ikke fordi bias kun findes i AI. Tværtimod. Bias findes også i menneskeskabt journalistik, i kilder, i redaktionelle traditioner og i journalistens egne antagelser.

Men AI kan gøre bias vanskeligere at opdage, fordi den pakker mønstre og skævheder ind i tilsyneladende neutralt og velformuleret sprog.

"*Jeg er bekymret lige nu, fordi folk ikke nødvendigvis er opmærksomme på bias. Vi er i en periode, hvor folk skal lære at forstå, hvad det er for teknologier, hvad de kan, hvad de ikke kan, og hvad farerne er,*" siger han. "*Men de er farlige i utrænede og uvidende hænder.*"

Samtidig ser han en mulighed. AI har gjort bias til et langt mere synligt emne. Og sprogmodeller kan faktisk også bruges til at lede efter bias i vinkler, citater og artikeludkast — hvis journalisten ved, hvad værktøjet kan og ikke kan.

Derfor mener han ikke, at AI kun skal forstås som en ny kilde til bias. Den kan også blive en anledning til, at journalister bliver mere opmærksomme på deres egne blinde vinkler. Men kun hvis de bliver trænet i at bruge den kritisk.

På DMJX skal de studerende bruge AI — men ikke til at springe læringen over
På DMJX er spørgsmålet ikke, om de studerende skal bruge AI. Det skal de. Men de skal lære at bruge teknologien uden at lade den overtage det, de selv skal lære.

"*Vi opfordrer dem til at bruge det rigtig meget. De må ikke skrive deres artikler, og de må ikke lave deres refleksionsrapport med AI. Men vi opfordrer dem til at bruge det som sparringspartner. Vi opfordrer dem til at være nysgerrige, eksperimentere og tale om det,*" siger Peder Hammerskov.

Han beskriver DMJX's tilgang som et forsøg på at skabe refleksion. De studerende skal ikke bare lære konkrete værktøjer. De skal have et sprog for, hvad AI er, hvad teknologien kan, hvad den ikke kan, og hvilke risici der følger med.

"*Det, der fylder meget for os, er at få skabt refleksionen. Hvis det skal lykkes, skal vi starte med at give dem AI-literacy — et fundament at stå på med viden om teknologien.*"

Peder Hammerskov arbejder derfor med AI-moduler, der kombinerer online læring med fysisk undervisning. De skal blandt andet handle om teknologiens muligheder og begrænsninger, ophavsret, faglighed, risici og journalistisk ansvar.

Men han understreger, at det vigtigste næsten ikke er selve værktøjsbrugen. Det vigtigste er, at de studerende bliver i stand til at tale kvalificeret om den.

"*Det vigtigste er næsten, at de begynder at få et sprog for at kunne tale om AI. Det gør dem forhåbentlig bedre til at reflektere over deres brug — både selv og når de skal skrive om det i deres reflektionsrapporter.*"

Mennesket først, mennesket sidst
I undervisningen forsøger DMJX ifølge Peder Hammerskov at arbejde med en tilgang, hvor mennesket både starter og afslutter processen.

"*Vi starter hos mennesket, der sætter rammerne. Så kan man tage det med over i AI og diskutere det med den. Og så kommer vi tilbage til mennesket til sidst, som verificerer, er kritisk, kvalificerer og holder øje.*"

Det er en model, der skal forhindre, at AI bliver en genvej uden læring. Først skal den studerende selv formulere intentionen, opgaven og kriterierne. Derefter kan AI bruges som sparringspartner. Til sidst skal den studerende tilbage og vurdere, hvad der kom ud af processen. Det lyder enkelt, men rummer et grundlæggende dilemma for journalistuddannelsen. De studerende skal lære at bruge et værktøj, der kan hjælpe dem. Men de skal samtidig opbygge den faglighed, der gør dem i stand til at vurdere, om værktøjet hjælper dem godt.

En erfaren journalist kan ofte se, når et AI-output er overfladisk, skævt eller faktuelt usikkert. En studerende er stadig ved at lære, hvad et godt interview, en stærk vinkel eller en holdbar artikelstruktur er.

"*Der er forskel på at arbejde som erfaren journalist og bruge AI — og så sidde på Journalisthøjskolen og skulle vurdere noget, der kommer ud af den, når man endnu ikke ved så meget om journalistik,*" siger han.

Derfor er AI i uddannelse ikke kun et spørgsmål om snyd. Det er et spørgsmål om læring.

"*Hvad er læring egentlig i en verden, hvor der er AI? Hvordan skaber man læring på trods af AI? Det er faktisk ret svært og ret interessant,*" siger han.

Den egentlige opgave er dømmekraft
Når Peder Hammerskov ser længere frem, tror han ikke nødvendigvis, at journalister forsvinder. Men rollen vil ændre sig. Journalisten kan i højere grad blive den, der definerer opgaven, samler dokumentation, finder kilder, vurderer vinkler, styrer AI-processen og tager ansvar for det færdige resultat.

"*Jeg tror allerede, journalistrollen er ved at ændre sig i retning af, at vi måske alle sammen bliver mere og mere redaktører,*" siger han. "*Jeg tror også, det bliver et parameter, man kan differentiere sig på — altså det her med, at man ikke bruger AI, og at man er et mere menneskeskabt produkt.*"

Men uanset hvor meget AI kommer til at fylde, bliver det afgørende spørgsmål ikke kun, om teknologien er brugt. Det bliver, hvordan den er brugt, og om medierne kan forklare det på en måde, brugerne forstår. Peder Hammerskov peger på et dilemma: Publikum vil gerne vide, hvis AI er brugt, men reagerer ofte negativt, når de får det at vide. Derfor kan transparens ikke reduceres til en mærkat ved hver eneste artikel. Medierne skal være åbne, når AI har haft afgørende betydning for journalistikken, men de skal også kunne forklare, hvad AI-brugen faktisk består i.

"*Jeg tror ikke, vi skal skrive ved hver eneste artikel, at vi har brugt AI. Men modsat skal vi heller ikke være utransparente omkring det. Det værste ville næsten være, hvis vi forsøgte at putte med det,*" siger han.

Den egentlige opgave er at sikre, at brugen bliver reflekteret. Det kræver regler, men ikke kun regler. Det kræver værktøjer, men ikke kun værktøjer. Det kræver undervisning, men ikke kun teknisk oplæring. Det kræver en journalistisk kultur, hvor man kan tale åbent om AI-brug, fejl, grænser, fristelser og ansvar. For den største risiko er ikke, at journalister bruger AI. Risikoen er, at de bruger den uden at forstå, hvad den gør ved deres proces, deres produkt og deres faglige dømmekraft. Og derfor er den svære del ikke at få journalister til at bruge AI. Det er at få dem til at bruge den reflekteret.

Illustration: Genereret med ChatGPT

TEMA: NÅR MASKINEN VÆLGER

KAN MAN MÅLE BIAS I MEDIER

Af Nicolai Hyllested og Mark Sinclair Fleeton

Bias er den skævhed, der opstår, når et system ikke bare gengiver verden, men vægter den. Det kan være åbenlyst, når noget fremstilles ensidigt eller stereotypt. Men det kan også være skjult: i tone, i vinkling, i kildevalg, i fravalg, i det, der virker neutralt, men som allerede trækker i en bestemt retning.

I journalistik har bias altid eksisteret. Men med AI bliver problemet mere vanskeligt, fordi skævheden ikke kun ligger i den enkelte journalists valg. Den kan også ligge i data, modeller, prompts, guardrails og de usynlige prioriteringer, der former en tekst, før den overhovedet når frem til læseren.

Derfor er spørgsmålet ikke kun, om AI i medierne er biased. Spørgsmålet er også, om medierne kan opdage det, dokumentere det og måle det. Nicolai Hyllested byggede en AI-måler for at finde skævheder i artikler. Men forsøget afslørede hurtigt et større problem: Der findes ingen objektiv baseline.

Den tekniske og den menneskelige bias

Bias er den skævhed, der opstår, når et system ikke bare gengiver verden, men systematisk vægter den på bestemte måder. I et AI-system kan det ske flere steder: i de data modellen er trænet på, i de instruktioner den får, i de sikkerhedslag og filtre der styrer dens svar, og i den måde output bliver rangeret, forkortet eller præsenteret på. Teknisk bias er derfor ikke kun et spørgsmål om, at modellen siger noget åbenlyst forkert eller diskriminerende. Det kan også være, at den oftere forbinder autoritet med bestemte typer personer, foreslår bestemte vinkler frem for andre, udelader perspektiver eller formulerer sig med en tone, der får én fortolkning til at fremstå mere naturlig end en anden. Bias er med andre ord ikke bare fejl i systemet. Det er mønstre i systemets prioriteringer.

Ofte bruger vi imidlertid ordet bias langt mere løst. I praksis kalder vi tit noget biased, når vi oplever, at en tekst, en vinkel eller en prioritering går imod vores egne holdninger eller verdensbillede. Bias bliver med andre ord ikke kun en teknisk eller analytisk kategori, men også en følelsesmæssig reaktion: en måde at sige, at noget føles skævt, urimeligt eller politisk farvet.

Det gør bias-begrebet vanskeligt, fordi vi som læsere og borgere sjældent møder tekster neutralt. Vi har allerede vores egne erfaringer, værdier og antagelser med ind i læsningen. Derfor kan det, den ene læser opfatter som nøgtern journalistik, af en anden opleves som ideologisk slagside. Og netop dér opstår det vanskelige spænd mellem menneskelig oplevelse og teknisk analyse: Bias er både noget, der kan være indlejret i systemer og

sproglige mønstre, og noget, vi som mennesker projicerer ind i tekster, når de rammer os på den forkerte side af vores egne overbevisninger.

Den objektive måler
Det store spørgsmål er selvfølgelig, hvordan man måler bias uden selv at blive en del af målingen og lægge sin egen bias ind i mellemrummet. Og den helt store elefant i rummet er selvfølgelig også: Hvordan måler vi overhovedet bias?

Bias er ikke et ord eller en sætning. Det er en vægtning. En skævvridning. Men i forhold til hvad?

Andre "målere" på markedet har forsøgt at løse det på forskellige kreative måder, med større eller mindre held. Kigger man lidt nærmere efter, viser det sig hurtigt, at de ofte virker i ret snævre områder, men ikke løser problemet generelt. Da Nicolai Hyllested begyndte at se på det, stod det hurtigt klart, at andre selvfølgelig også havde fået den tanke, at der lå et reelt problem her. Bagefter pegede forskellige eksisterende løsninger i samme retning. Og ja, de kan være brugbare. Men de bygger som regel på menneskelige vurderinger, politiske akser eller på forhånd definerede kategorier.

Det vil sige, at mange biasmålere i praksis først indfører et perspektiv og derefter måler ud fra det. For eksempel ved at definere en observatør eller trække en højre-venstre-akse ned over et helt medie. Nicolai Hyllested gik en helt anden vej, mere ved et tilfælde end ved stor visdom.

I et helt urelateret emne omkring bias havde han brug for at kunne få 100 procent ud af tre biasspor, altså lægge dem sammen, og den eneste måde, han kendte til, var vektorer. Da først den model var på plads, kunne bias operationaliseres som et tal, i stedet for bare at være en fornemmelse i teksten, og følges gennem systemet for at se, hvor meget det drejede sig i den anden ende.

Så var opgaven "bare" at isolere, hvad der skyldtes bias, og hvad der skyldtes tone, konvergens og alt det andet, der forstyrrer målingen. Millioner af tokens senere var næste skridt at overføre det til et program, der kunne måle bias i medierne. Det var i hvert fald ambitionen.

Det viste sig hurtigt, at når man går fra et system, hvor man kan kontrollere for næsten alt, til den virkelige verden, så sker der et kompleksitetsskifte, der er fuldstændig vanvittigt. Verden er ikke lineær. Og den er i hvert fald ikke ren nok til, at man bare kan bygge en baseline og måle op imod den.

Den umulige baseline
Det største problem er uden tvivl at skabe en baseline, og de studier, der siden er læst, viser også, at Nicolai Hyllested på ingen måde er den første, der er kørt fast dér. Man bliver nødt til at skabe et ståsted, hvorfra man kan se sin vektor dreje. Og her ligger der en stor fare: at man lægger sin baseline så højt, at alle andre målinger kommer til at se flade ud, fordi de ender på niveau. Hvis alle bjerge har samme højde, inklusive ens eget, er det pludselig svært at se, hvilket der faktisk er højest.

Derudover er der helt simple problemer som sample size og selve udvalget, man bruger til at beregne sin baseline ud fra. Og hvis man ikke tidligere har arbejdet med dataindsamling eller statistik, er det måske meget fair, at man ikke når helt derhen, når man bygger sin måler.

Men det helt grundlæggende problem er mere irriterende end som så: Hvordan finder man neutrale medier at måle fra? Medier der er fair, præsenterer begge sider nogenlunde lige, og gør det i stor nok skala til, at man faktisk kan bruge dem til noget. Det var dér, forsøget stødte på sit største problem. For i praksis er det tæt på umuligt. Ikke fordi der ikke findes relativt nøgterne medier, men fordi "neutral" i praksis ikke er en ren og objektiv kategori.

Så valgte Nicolai Hyllested en anden tilgang. Den historie, der skulle undersøges, blev hentet ind, fundet hos hovedkilder, der på forhånd var defineret, for eksempel Reuters, og derfra blev der bygget et cluster rundt om den samme historie. Samtidig blev der brugt et par andre greb. For eksempel blev historien strippet ned til sin grundform og lagt ind som et ekstra kontrollag i baselinen.

Det viser sig desværre hurtigt, at i en verden af bias og medier er det ikke helt fair kun at tale om bias. Og det er her, man mister fodfæstet. Ja, et medie kan godt være biased eller mindre biased. Men hvis tonen er hård, hvis den ene side fremstilles upræcist, hvis den ene part fylder mere, eller hvis noget vigtigt er udeladt, så kan en historie komme til at fremstå biased uden nødvendigvis at være det på en simpel politisk akse.

Så står man pludselig ikke bare med ét problem, men med flere lag oven i hinanden. Flere tal. Flere steder, hvor en læser kan springe over, hvor systemet kan skævvride, og hvor en historie kan blive båret videre som bias til en anden persons egen agenda.

Vores verden er bias
Bias bliver aldrig noget, man bare kan trække ud i lyset og derefter være færdig med. Det er snarere et grundvilkår, man er nødt til at arbejde systematisk med. Verden er fuld af skævheder, interesser, perspektiver og indbyggede prioriteringer, og journalistikken står ikke udenfor det. AI gør ikke problemet mindre. Den gør det snarere mere komplekst, fordi de redaktionelle valg nu i stigende grad flettes sammen med tekniske valg, som ikke altid er synlige for hverken redaktion eller læser.

Risikoen er derfor ikke nødvendigvis, at AI gør journalistikken mere åbenlyst skæv. Risikoen er, at skævheden bliver mere ensartet, hurtigere og lettere at skalere. En AI følger den retning, den er sat i, og dermed kan et medies eksisterende linje blive forstærket, uden at det nødvendigvis opdages undervejs. Det gør ikke journalistikken mere neutral. Det gør bare dens mønstre mere stabile.

Bias skal kunne spores
Det efterlader medierne et sted mellem erkendelse og ansvar. De kan næppe bygge en endelig og objektiv biasmåler, der én gang for alle afgør, om en tekst er neutral eller skæv. Men de kan godt blive langt bedre til at gøre deres egne valg synlige, teste deres systemer, sammenligne output, føre log over brugen af AI og diskutere, hvilke mønstre de selv er med til at reproducere.

Målet er ikke at rense journalistikken for enhver skævhed - det har aldrig været muligt - men at gøre bias mindre skjult, mindre automatisk og mindre uimodsagt. Hvis AI i stigende grad bliver en del af journalistikken, er det måske ikke den perfekte måler, medierne har mest brug for, men en mere ærlig praksis: en, der erkender, at også maskinens neutralitet er et redaktionelt valg.

Læs Nicolai Hyllesteds papers om hans arbejde på hans LinkedIn-profil: https://www.linkedin.com/in/nicolai-hyllested-86849976/

Illustration: Genereret med ChatGPT

BIAS ER IKKE PROBLEMET. DET ER DEN SKJULTE BIAS, DER ER

- EN PERSONLIG REFLEKSION OVER LLM'ER, SKJULT VÆGTNING OG HVORFOR BIAS BLIVER FARLIGERE, NÅR DEN OPFØRER SIG SOM NEUTRAL HJÆLP

Af Nicolai Hyllested

Jeg elsker bias. Den der ubestemmelige størrelse, som alle slynger efter hinanden og bruger til at slå andre oven i hovedet med, mens de selv naturligvis står skinnende rene og objektive tilbage.

Lad os starte et personligt sted. Ikke alt for dybt, bare nok til at være irriterende ærligt.
I mange år var jeg overbevist om, at jeg selv var objektiviteten med ben. Jeg stod så længe på logikkens alter, at rationalet nærmest føltes som min religion. Der skulle både et forhold og en personlig krise til, før det gik op for mig, at jeg selvfølgelig heller ikke var fri for bias.

Så ramte AI-verdenen. Og pludselig var bias ikke bare noget, vi skulle tage stilling til i medier eller politik. Det var noget, vi sad og talte direkte med. Noget der svarede igen. Noget der ikke bare skrev tekst, men opførte sig som et spejlbillede af os selv.

Vi har lært at tale om bias i aviser, på tv og i den offentlige debat. Fint nok. Men med LLM-systemer er vi nået til et andet sted. Her bliver bias ikke bare læst. Det bliver oplevet. Det bliver pakket ind i samtale, tone, hjælpsomhed og genkendelighed. Og jo mere de her systemer bliver normale, jo mere kommer de også til at påvirke os, langt mere end mange har lyst til at indrømme.

Når de kloge stiller sig op og forklarer AI, starter de næsten altid samme sted: træningsdata, vægte og sandsynligheder. Og ja, det betyder noget. Men for brugeren er det sjældent dér, slaget står.

Vi tror, at bias først opstår i det øjeblik, vi får et svar. Men allerede før dit spørgsmål når frem til modellen, er der truffet valg på dine vegne. Det spørgsmål, du tror, du stiller, er derfor ikke nødvendigvis det spørgsmål, modellen faktisk svarer på.

I den anden ende bliver svaret også afgrænset, blandt andet af noget så lavpraktisk som længde, tokenbudget og hvad der passer ind i brugerfladen. Også det er en vægtning. Også det former det svar, du ender med at læse.

Og mellem de to punkter ligger der flere lag: guardrails, runtime-logik, systemstyring og andre indgreb, som er med til at formulere det færdige svar. Læg dertil tone, alignment, konvergens og det dybere indre maskinrum som residual stream, så begynder pointen at blive ubehageligt klar: Vi ser et output, men vi aner kun i begrænset grad, hvilke skub, filtre og prioriteringer der har formet det undervejs.

Min påstand er derfor ret enkel. I praksis er der fire ting, der former vores oplevelse af AI langt mere direkte end de fleste forklaringer om vægte og sandsynligheder: bias, tone, alignment og konvergens.

Tone, alignment og konvergens kan i et vist omfang styres. De kan justeres, dæmpes, trimmes og pakkes pænt ind. Men at der findes noget i et system, nemlig bias, som ingen reelt har styr på, burde gøre alle en smule urolige.

Og her mener jeg ikke bare den åbenlyse bias, som alle straks leder efter. Vi er vant til at råbe biased i kvinde- og mandespørgsmål, eller når en indvandrer prøver at komme ind på det lokale diskotek. Og helt ærligt, der er LLM-modellerne ofte blevet ret gode. Sikkert fordi det er et fokusområde, og de store firmaer udmærket ved, at det er noget af det første, brugerne tester. Det ville faktisk være mærkeligt, hvis kommercielle firmaer med etiske ambitioner ikke havde styr på netop det lag.

Det mere bekymrende er noget andet. Der ser ud til at være domæner, hvor der optræder en mere strukturel bias. Jeg bruger ordet strukturel her, selv om det teknisk set er lidt mere nuanceret end som så. Men ordet er brugbart nok til formålet. Pointen er, at der findes skævheder, som ikke bare forsvinder, fordi man lægger guardrails og runtime-lag ovenpå. Oversat til almindeligt dansk: der findes bias, som dem der bygger systemet, ikke bare kan fikse med lidt pæn kodelogik og et sæt regler ovenpå modellen.

Vi kan alle være nogenlunde enige om, at vægtene har deres egne indbyrdes forhold, og at vi som brugere ikke har en jordisk chance for at kende dem. Helt ærligt virker det nogle gange heller ikke som om, dem der bygger systemerne, forstår mere end stykvis af det, de står med. Mere bekymrende er det måske, at måden man fodrer rådata ind i systemet på, også har væsentlig indvirkning på, hvordan vægtenes indbyrdes forhold udvikler sig, og dermed på små forskelle i bias i den endelige model.

I en verden af residual stream, residual reasoning og jagten på det næste spring i compute er det ikke mærkeligt, at bias bliver overset.
Forskningen jagter fremdrift. Firmaerne jagter kapabilitet. Brugerne tester de mest åbenlyse skævheder først: mand, kvinde, chef, ansat, navn, hudfarve, politiske markører. Det giver mening. Det er det synlige. Det er det, man kan pege på.

Men hvis vi kun ser på, hvad bias betyder i de oplagte cases, og ikke på hvilken adfærd en LLM faktisk har på tværs af domæner, så skyder vi os selv i foden.

For lige nu vurderer vi ofte LLM'er lidt som vi vurderer sociale medier, ud fra det synlige output. Det er en fejl. Output er ikke en særlig god benchmark for det, vi ikke ser. Og slet ikke for det, der sniger sig ind under radaren.

Bias er i sig selv ikke et problem. Bias unchecked og uadresseret er.

Vi er som mennesker vant til at være omgivet af bias overalt. I andre mennesker. I bøger. I medier. I kulturen generelt. Vi ved, uden at nogen behøver sige det højt, at der er en vægtning. At nogen ser verden fra et sted. Det er først, når denne bias ledes ind ad bagdøren, uden om vores forsvarsværker og vores kritiske sans, at problemerne for alvor opstår.

Vi anerkender uden større diskussion, at en kultur hvor kvinder konstant omtales som det svage køn, ikke er sund. Men hvad med en hr-afdeling, der bruger AI uden at vide, at status og autoritet er problembørn? Hvad med et system, der lægger skjult vægt på kompetence på en måde, ingen i rummet opdager, fordi alle regner med, at OpenAI eller Anthropic selvfølgelig har styr på det?

Ingen forventer den spanske inkvisation. Nej. Men i det mindste vidste man, at den fandtes. Her er vi på bar bund. Og hvis bias faktisk er strukturel, så har de jo af gode grunde heller ikke fuld kontrol over den.

Vi er på vej ind i en tid, der sætter vores tillid endnu mere på prøve. I den digitale tidsalder havde vi trods alt et flow, der i det mindste var lineært og forståeligt for nogen mennesker på kloden, uanset hvor komplekst det blev. Nu træder vi ind i et univers, hvor ingen reelt ved, hvordan systemet faktisk virker.

Det univers minder mindre om klassisk software og mere om en hjerne. Vi ved, at der sker noget. Vi kan se et udfald. Men de enkelte processer er grundlæggende lukket land og kan kun tilnærmes gennem modeller, målinger og kvalificerede gæt.

Og her er bias væsentlig. For bias er noget af det eneste, der rammer os direkte, uden at vi eller andre nødvendigvis kan skrue ret meget på den underliggende sammenhæng mellem vægtene. Det er også derfor, det er så vigtigt at forstå, at vi ikke står neutralt i mødet med de her systemer.

Jeg siger ikke, at vi skal være bange for AI. Snarere det modsatte. Vi skal bruge det. Vi skal lære det. Vi skal presse det. Men vi skal også omgås det, som vi ville omgås et andet menneske. Med åbenhed, ja. Men også med paraderne oppe.

Vi må ikke lade os lulle i søvn og tro, at vi står upåvirkede i mødet med noget, der kan blæse lige gennem vores kognitive forsvar, før vi overhovedet opdager, at vi burde have haft dem oppe.

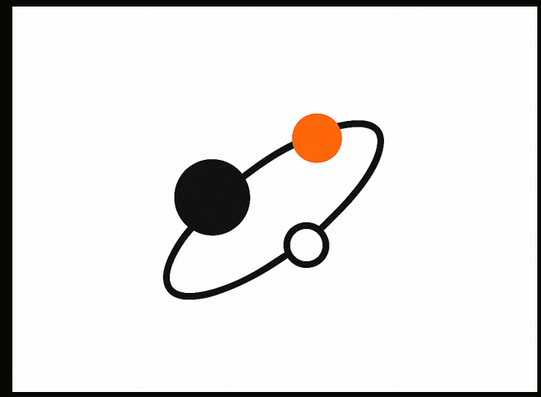
Skal vi så helt opgive de her systemer?

Nej. Det er egentlig ikke min pointe.
Min pointe er, at vi står et sted, hvor vi har stået før med ny teknologi, i ukendt territorium. Noget, der kommer til at påvirke os på måder, vi endnu ikke forstår fuldt ud.

Bare se på sociale medier. De kom ikke med en seddel, hvor der stod, hvad de ville gøre ved vores opmærksomhed, vores adfærd og vores kultur. For mange er det gået fint nok. For andre har konsekvenserne været ret katastrofale.

Hvis vi allerede tidligt kan få klarhed over de problemer, der findes i LLM-systemer, har vi i det mindste en chance for at komme nogle af dem i forkøbet.

Vi redder ikke alle fra at falde i et hul. Men måske kan vi gøre landingen en smule mindre brutal.



TEMA: NÅR MASKINEN VÆLGER

PROJEKT Y VIL IKKE BARE LAVE ET MEDIE - DE VIL EJE MASKINEN BAG

AF MARK SINCLAIR FLEETON

Illustration: Genereret med ChatGPT

Det nye ved Projekt Y er ikke, at det er borgerligt, eller at det bruger AI. Det nye er forsøget på at samle journalistik, teknologi og forretning i én konstruktion – hvor mediet skal fungere som proof of concept for en platform, der kan sælges videre.

Mediet er en synlig del af noget større
Bag Projekt Y sidder en kreds af investorer med baggrund i finans og tech: Lars Seier Christensen, Lars Tvede fra Supertrends og Morten Wagner er blandt de 12-15 aktionærer, projektet er bygget op om. Torben René Larsen er CEO, Michael Dyrby er chefredaktør. Projektet er organiseret som et ApS med decentral struktur og folk spredt fra København til Schweiz.

Det, der skiller projektet ud, er ikke, hvem de er, men hvad de vil bygge. Ikke bare et nyhedsmedie, men en sammenhængende konstruktion, hvor workflow, produktion, distribution og licenserbar teknologi hænger sammen – og hvor mediet er den synlige del af noget større.

"Så der er to. Der er Y-Rails og Y.dk," siger Torben René Larsen.

Y.dk er mediet. Y-Rails er platformen bag. Det ene skal udkomme med journalistik. Det andet skal kunne licenseres til andre som Y-Label eller Y-Tools. Dermed er Projekt Y ikke bare historien om et nyt medieprojekt. Det er historien om et forsøg på at bygge medieinfrastruktur som forretning.

Et borgerligt projekt, der blev til et AI-projekt
Idéen begyndte før AI-hypen. Ifølge Torben René Larsen havde gruppen bag i flere år talt om at starte et nyt medie, fordi de oplevede det danske medielandskab som for smalt og for præget af dagsordener.

"Vi vil gerne lave noget mere faktabaseret. Og så vil vi jo selvfølgelig også gerne, da vi er et borgerligt, liberalt land, konservativt eller hvad endnu efterhånden, men i hvert fald et borgerligt orienteret land, så skulle det også have en eller anden borgerlig impact i samfundet," siger Torben René Larsen.

Så kom de generative AI-værktøjer. Det ændrede projektets karakter. Det, der begyndte som ønsket om et nyt medie, blev samtidig et forsøg på at bygge det 'fra grunden af som et AI-projekt'.

Michael Dyrby blev hentet ind som chefredaktør. Rundt om ham blev der samlet en teknisk kreds med folk fra blandt andre Supertrends, SpeakerBee og B-Stream. En bestyrelse er endnu ikke på plads, men er ifølge Larsen under etablering.

Allerede dér bliver det tydeligt, at Projekt Y ikke er bygget som en klassisk redaktion. Det er bygget som et projekt, hvor journalistik og teknisk udvikling vokser frem side om side.

Ikke bare et medie - også et techfirma
Folkene bag siger det selv ret direkte.
"Så hvad er vi? Er vi et techfirma, eller er vi et mediefirma? Ja, vi er nok et tech-mediefirma. Fordi de to ting hænger sammen," siger Michael Dyrby.

Torben René Larsen skærer det endnu hårdere til. Ser man på projektet lige nu, er der flere it-folk end journalister. *"Så lige nu er det jo mere et IT-firma."*

Sammenligningen, de selv trækker frem, er Saxo Bank. En virksomhed der begyndte med at bygge en handelsplatform til egne kunder, fandt ud af at platformen kunne sælges videre – og endte med at være et IT-firma, der tilfældigvis også er bank. Analogien er ikke tilfældig. Den beskriver præcist, hvordan Projekt Y tænker om sig selv.

De taler ikke om AI som et ekstra redskab til journalister. De taler om et samlet workflow. Jens Helstrup, teknisk profil med base hos Supertrends i Schweiz, beskriver en platform der monitorerer data, gør ustruktureret input struktureret og sender det videre i et generativt system, som kan producere tekst, lyd og video.

"En god prompt er ikke nok," siger han.

Det er en vigtig del af forståelsen af projektet, fordi den markerer forskellen til meget af den AI-brug, etablerede medier taler om. Der bruger man værktøjer i kanten af den eksisterende produktion. Projekt Y vil bygge hele kæden om.

Dyrby beskriver det som et system, hvor AI er inde 'fra idé og data til udgivelse og så et loop tilbage'. Det er ikke bare endnu et tool. Det er arkitekturen.

Mediet som ansigt, platformen som forretning
Projektets todeling er enkel nok at forstå. Y.dk skal være det synlige medie. Y-Rails skal være maskinen bag. Men den todeling er også nøglen til at forstå den økonomiske idé.

Projekt Y siger åbent, at de ikke tror på, at et dansk medie i sig selv bliver den store forretning. Mediet skal på sigt tjene penge på annoncer og abonnementer, 'noget ligesom alle andre'. Men det er ikke dér, de selv ser den store værdi.

"Vi tror mest på platformen i forhold til at skabe finansiel værdi," siger Torben René Larsen.

Det er et afgørende udsagn. For hvis han har ret, er journalistikken ikke kun målet. Den er også demonstrationen af, at systemet virker.

Det billede bliver kun skarpere af de opfølgende svar. Y-Rails har et par testkunder, som Projekt Y ikke vil navngive, men endnu ingen betalende kunder. Y-Label og Y-Tools er ikke færdige produkter i markedet endnu. Førsteprioriteten er at få Y.dk i luften, så mediet kan fungere som et live proof of concept.

Det er svært at læse på anden måde: Y.dk skal ikke bare være et medie. Det skal også være showroom.

Human in the loop – og ambitionen om at slippe det
Projekt Y placerer sig foreløbig i et mellemstadie mellem klassisk journalistik og fuld automation. Mere automatiseret end de fleste eksisterende mediers AI-brug, men stadig ikke fuldt autonomt.

Projektet arbejder med en intern trappemodel for automatisering. I den nederste ende ligger klassisk journalistik uden AI. Derefter kommer AI som hjælpemiddel, AI som producent under redaktionel kontrol, AI-drevet publicering med redaktionel indgriben, og i toppen et fuldt autonomt niveau – det, de kalder L5.

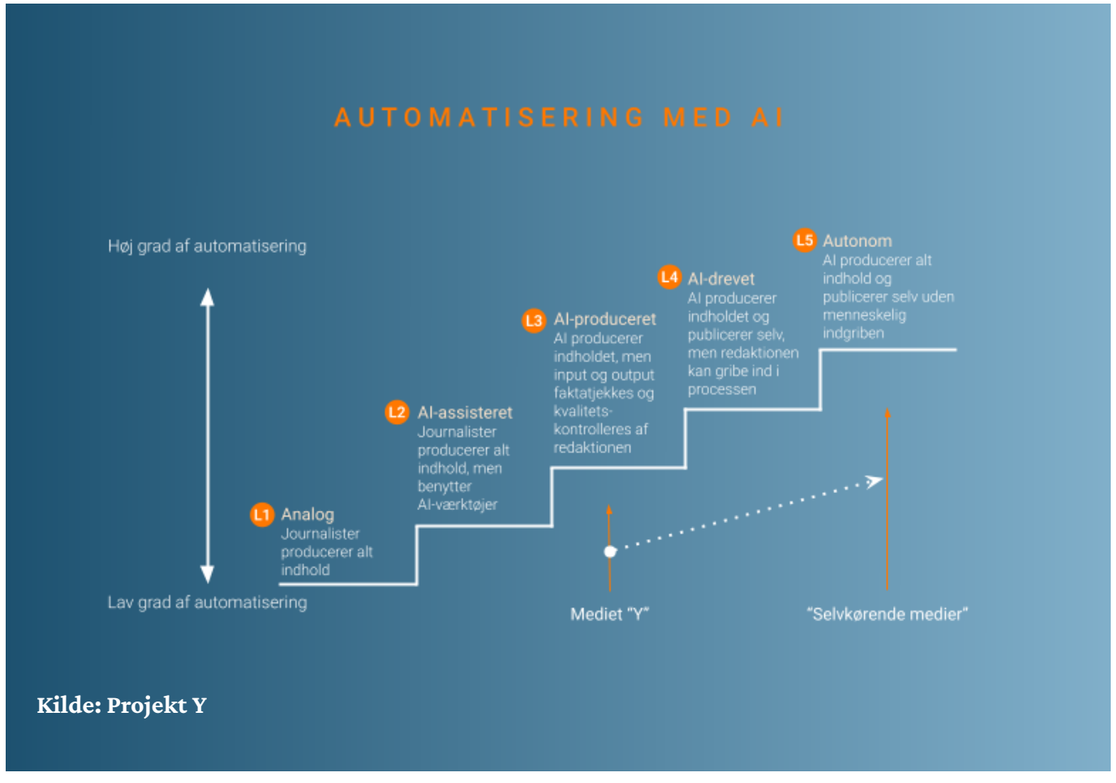
"Y-Rails vil arbejde på at flytte det helt op i L5," siger Michael Dyrby.

Ambitionen er ikke, at det menneskelige led skal være uforanderligt. Sigtekornet er niveau L4 – dér, hvor agenten overtager redaktørrollen.

"Det er der, hvor agenten i praksis bevæger sig ind i en redaktørrolle. Det er det, vi sigter imod. Lige nu er vi nødt til at have en eller anden form for redaktion. På et tidspunkt vil agenten overtage den rolle, som redaktionen eller redaktøren har i dag," siger Torben René Larsen.

Det er ikke noget, de taler udenom. Det er en del af projektets retning.

Det gør dog ikke Projekt Y til et rent førerløst medieprojekt fra dag ét. Tværtimod rekrutterer de nu redaktører – et aktuelt jobopslag viser, at de søger folk til det AI-drevne nyhedsmedie. Michael Dyrby beskriver redaktøren som en slags dirigent: en person, der vælger historie, sætter vinkel, får systemet til at producere, vurderer output og står med ansvaret. Det er altså ikke en klassisk redaktion, de forestiller sig. Det er en redaktion bygget til at styre en maskine – på vej mod at gøre sig selv overflødig.



Kilde: Projekt Y

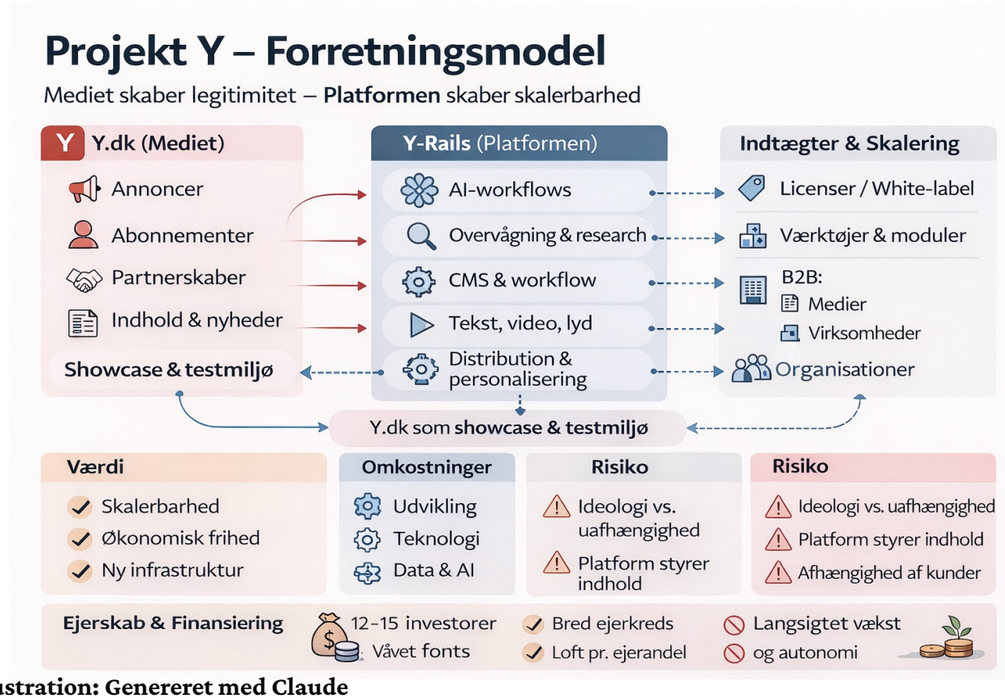


Illustration: Genereret med Claude

Bias, kilder og kontrol
Et af de mest sårbare punkter i hele projektet er spørgsmålet om bias. Det er også et af de mest komplekse – fordi det ikke er ét spørgsmål, men to. Det første handler om ideologisk bias. Og her er Projekt Y usædvanligt åbne. Michael Dyrby afviser ikke, at mediet har et udgangspunkt.

"*Bias er jo en interessant ting, fordi der er jo bias i alt, hvad der udkommer*," siger han.

Det er ikke en afvisning. Det er en vedkendelse. Projekt Y skjuler ikke, at det arbejder ud fra et borgerligt værdigrundlag – de gør det tværtimod til et bevidst valg, at brugerne skal vide, hvad de får. På det punkt er de mere transparente end mange etablerede medier.

Men det er det andet spørgsmål, der er sværere: den systemiske bias, der kan opstå, når AI-modeller producerer i stor skala på baggrund af træningsdata, promptdesign og kildekuratering. Her svarer Jens Helstrup på procesplan: systemet bygger på referencekilder, promptstyring og brug af flere modeller i samspil, der kan kontrollere og korrigere hinandens output. Kildevalidering sker, inden materiale overhovedet kommer ind i systemet.

Det er mekanismer, men det er ikke målemetoder.
Projekt Y kan endnu ikke dokumentere, hvad systemet faktisk producerer i praksis – fordi det ikke er i drift. Den egentlige test af, om processerne holder, kommer først, når mediet er oppe at køre, og output kan måles og efterprøves. At have checks er ikke det samme som at have bevis for, at de virker.

Det fjerner ikke ambitionerne. Men det forskyder spørgsmålet: fra om Projekt Y *har* tænkt over bias, til om de *kan* eftervise, at de håndterer den.

Ejerskab: bred kreds, bundet ramme
Projekt Y er finansieret af 12-15 investorer, og strukturen er bevidst designet til at forhindre, at nogen kan skyde sig til en dominerende position. Torben René Larsen fortæller, at der er sat et loft over, hvor meget den enkelte kan eje.

Det lyder som en sikring af uafhængighed. Men her er en væsentlig forskel, som Larsen oplyser i et opfølgende svar: loftet over ejerandele er ikke skrevet ind i vedtægterne. Det er i stedet skrevet ind i ejeraftalen – den aftale, som nye medejere skal tilslutte sig, og som fastlægger mediets værdier og retning.

Det er ikke det samme. En vedtægtsmæssig begrænsning er formel og juridisk sværere at ændre. En ejeraftale er en aftale mellem de nuværende parter – og kan potentielt forhandles, omskrives eller udvandes, efterhånden som investorsammensætningen ændrer sig.

Projekt Y forsøger altså ikke at sikre sin uafhængighed gennem hård, formel strukturering. Det forsøger at sikre den gennem en bredere ejerkreds og et værdifællesskab, der er skrevet ind fra starten.

Det kan læses som et pragmatisk og fleksibelt alternativ til stiftelsesfondsmodellen. Men det kan også læses som noget andet: en konstruktion, hvor det borgerlige værdigrundlag ikke er en redaktionel linje, der kan diskuteres, men et kontraktuelt bindende udgangspunkt.

Det er dér Projekt Y's egentlige spænding ligger. Projektet peger mod en medieform, der måske kan blive mere økonomisk selvbærende. Men det gør det ikke neutralt. Værdigrundlaget er skrevet ind i ejeraftalen.

2026 – hvis kvaliteten holder
Projekt Y er ikke i luften endnu. Men Michael Dyrby siger, at Y.dk kommer i 2026. Og det kommer ikke 'med hele butikken til at starte med'.

"Her skal du faktisk komme sidst med den rigtige teknologi," siger Torben René Larsen.

Det er en god sætning, fordi den siger noget præcist om deres egen selvforståelse. De ser ikke feltet som et kapløb om at komme først. De ser det som et kapløb om at gøre det rigtigt.

Saxo Bank-analogien vender tilbage her: Saxo Bank brugte år på at bygge platformen, inden den gav mening at sælge. Projekt Y tænker tilsyneladende på samme måde om Y-Rails. Mediet er første trin. Platformen er det egentlige mål.

Det egentlige spørgsmål
Projekt Y er endnu ikke et medie i drift. De har ingen betalende kunder på platformsiden. Bestyrelsen er ikke på plads. Mange af de store ord er stadig fremtidsord. Men det gør ikke projektet mindre interessant. Tværtimod.

Det interessante er, at Projekt Y siger højt det, mange andre kun eksperimenterer med i bidder: at AI ikke bare kan være et redskab i redaktionen, men selve infrastrukturen bag indhold, distribution, personalisering og forretning.

Dermed er det egentlige spørgsmål heller ikke kun, om Projekt Y lykkes. Det er, om de allerede nu peger på en medieform, hvor journalistikken ikke længere er hele forretningen, men én funktion i et større teknologisk system. Og hvis det er dér udviklingen bevæger sig hen, bliver den afgørende kamp i mediebranchen ikke kun om indhold, troværdighed eller ideologi. Den bliver også om, hvem der bygger, ejer og kontrollerer maskinen bag.

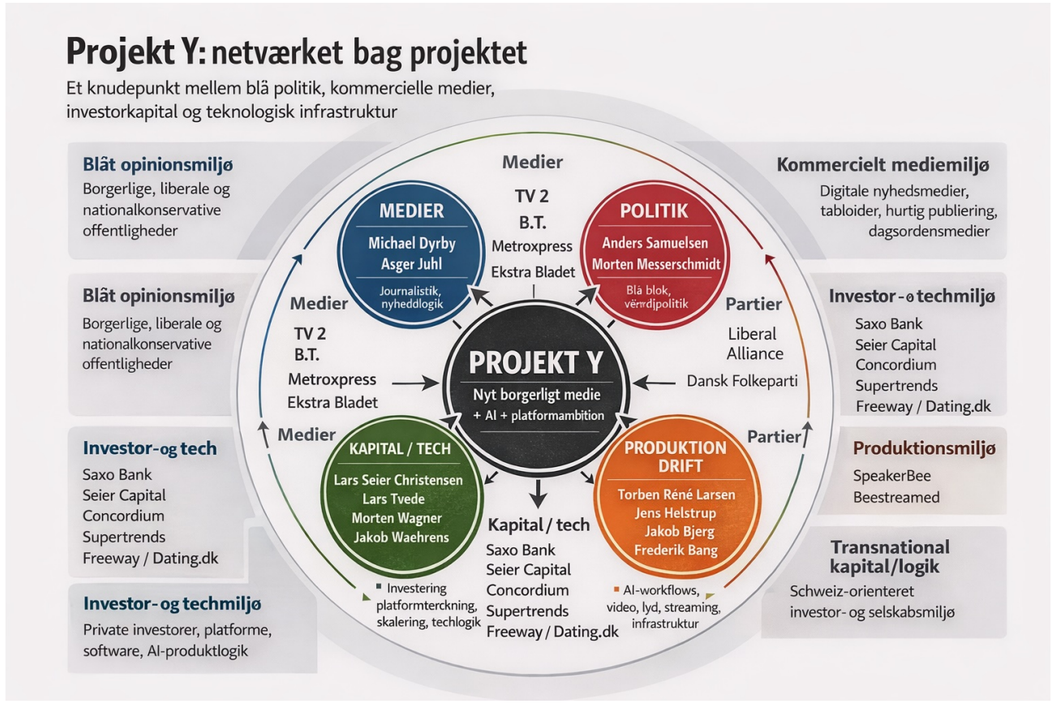


Illustration: Genereret med Claude

Illustration: Just Kjærgaard Pedersen

TEMA: NÅR MASKINEN VÆLGER

BIAS I AI ER IKKE DET, DU TROR

Af Just Kjærgård Pedersen, lektor ved Erhvervsakademi København

Bias er blevet et af de ord, som næsten automatisk dukker op, når man taler om kunstig intelligens. Hvis en chatbot svarer stereotypt, hvis en billedgenerator fremstiller mennesker på mærkelige måder, eller hvis et system tilsyneladende favoriserer bestemte kilder, perspektiver eller konklusioner, så siger vi hurtigt, at der er bias i systemet. Ordet bruges efterhånden så ofte, at det næsten lyder selvforklarende. Men det er det ikke.

For hvad betyder det egentlig, når vi kalder noget bias? Umiddelbart lyder svaret enkelt. Bias kommer fra det græske ord for skævhed. Noget der hælder i en bestemt retning. Noget er ikke neutralt. Men allerede her opstår det første problem. Hvis du siger, at noget er skævt, så må du vel også have en forestilling om, hvordan det ville se ud, hvis det ikke var skævt. Når vi på godt nydansk kalder et svar, en udvælgelse eller en vurdering biased, så sammenligner vi det, bevidst eller ubevidst, med en idé om noget mere retvisende, mere balanceret eller mere korrekt.

Det gør bias til et mere filosofisk begreb, end det ofte får lov til at være. For måske siger ordet noget grundlæggende om din egen forståelse af viden og virkelighed. Hvornår giver det mening at tale om bias? Og hvornår gør det ikke? Hvis du undersøger noget, hvor der faktisk findes en ekstern virkelighed, som vi kan ramme mere eller mindre præcist, så giver bias-begrebet umiddelbart god mening. Hvis du derimod bevæger os ind på spørgsmål om politik, moral eller det gode liv, bliver sagen mere uklar. Her er der sjældent en neutral facitliste, som en algoritme eller et menneske bare kan afvige fra. I sådanne tilfælde er det måske mere ærligt at tale om perspektiv, værdier og ståsted end om bias.

Det er netop denne sondring, der gør diskussionen interessant. I debatten om AI bruges bias ofte, som om ordet dækker det hele. Nogle gange handler det om fejl i forhold til en målbar virkelighed. Andre gange handler det om skæv udvælgelse. Og i andre tilfælde bruges ordet om noget, der snarere burde beskrives som en normativ eller politisk uenighed om, hvordan verden bør forstås. Når vi diskuterer bias i AI, diskuterer vi derfor hvad vi selv mener, at god viden er.

Hvad bias normalt betyder

Bias er bestemt ikke noget, som kun findes i kunstig intelligens. Tværtimod er det en kendt udfordring i al analyse, forskning og beslutningstagning. Hvis du interviewer tre af dine egne venner om et emne og bagefter siger, at du nu ved, hvad "folk" mener, så vil de fleste straks kunne se problemet. Det er ikke nødvendigvis fordi dine venner lyver, og heller ikke fordi dine spørgsmål er meningsløse. Problemet er, at udvælgelsen af respondenter allerede hælder i en

bestemt retning. Din viden er blevet formet af, hvem du har spurgt og hvem de er for dig.

Det samme gælder måden, du spørger på. Hvis du stiller forskellige mennesker forskellige spørgsmål, eller hvis dine spørgsmål er ledende, uklare eller bygget op på en måde, der inviterer til bestemte svar, så opstår der også en skævhed. Her ligger bias altså ikke i personerne, men i selve metoden. På samme måde kan bias opstå i fortolkningen. To personer kan godt sidde med det samme materiale og alligevel lægge vægt på forskellige dele af det. Den ene hæfter sig ved mønstre, der bekræfter en forventning. Den anden ser noget andet. Dermed bliver bias ikke kun et spørgsmål om data, men også om udvælgelse, spørgsmål, metode og læsning.

Det er netop derfor, bias er et så nyttigt begreb i mange sammenhænge. Det hjælper dig med at få øje på, at viden ikke bare opstår af sig selv. Den bliver frembragt gennem en proces, og den proces kan hælde. Nogle gange sker det helt åbenlyst. Andre gange er det langt mere subtilt. Et datasæt kan være systematisk skævt, uden at nogen bevidst har ønsket det. En interviewer kan påvirke svarene, uden at det er meningen. En analytiker kan se det, der passer bedst til en eksisterende antagelse, uden selv at opdage det. Bias er bestemt ikke det samme som ond vilje, men ofte bare et tegn på, at menneskelig erkendelse og vidensproduktion altid foregår gennem valg.

Når begrebet alligevel bliver mere vanskeligt, er det fordi forskellige typer skævhed ofte blandes sammen. Nogle former for bias handler om repræsentation. Har du spurgt de rigtige mennesker, eller kun dem, der ligner dig selv? Andre handler om måling. Har du målt det, du tror, du måler, eller har din metode skubbet resultatet i en bestemt retning? Andre igen handler om fortolkning. Ser du faktisk materialet, eller ser du mest det, du på forhånd forventede at finde? I den forstand er bias ikke én ting, men en samlebetegnelse for flere beslægtede problemer i produktionen af viden.

Når man taler om bias i AI, handler det sjældent kun om, at en model siger noget "forkert". Det kan lige så godt handle om, hvilke data modellen er trænet på, hvilke mønstre den derfor har lært, hvilke kilder den typisk fremhæver, og hvilke typer svar den lettere falder tilbage på. Som jeg bemærker i min guide til prompting, bliver varians ofte overset eller misforstået af nye brugere af kunstig intelligens. Men netop fordi AI både kan være variabel og skæv, bliver det vigtigt at skelne mellem forskellige typer problemer. Ikke alt, hvad der virker mærkeligt, er nødvendigvis bias. Og ikke al bias er af samme slags.

Men her opstår et mere grundlæggende spørgsmål. Når du siger, at noget er skævt, hvad er det så skævt i forhold til? Hvad er det normale, det neutrale eller det mere rigtige, som skævheden afviger fra? Det spørgsmål kan man ikke besvare alene ved at pege på dårlige data, upræcise spørgsmål eller uheldige prompts. Man er også nødt til at stoppe op og spørge, hvad viden egentlig er, og hvordan mennesker overhovedet kommer frem til den. Her støder man hurtigt på to meget forskellige måder at forstå viden på. Den ene tager udgangspunkt i, at der findes en verden derude, som vi kan beskrive mere eller mindre korrekt. Den anden tager udgangspunkt i, at vores viden altid vil være præget af sprog, perspektiv og den sammenhæng, vi befinder os i. De to måder at tænke på kaldes normalt positivisme og konstruktivisme, og de er relevante her, fordi de giver to forskellige svar på, hvornår det giver mening at tale om bias.

Når virkeligheden har et facit

Positivismen bygger på forestillingen om, at der findes en virkelighed uden for os selv, og at vi i princippet kan beskrive den mere eller mindre præcist. Verden er der, også når vi misforstår den. Fakta afhænger ikke af vores følelser.

Temperaturen udenfor er den temperatur, den er. Antallet af mennesker i et lokale er det antal, det er. Hvis du måler forkert, har du ikke skabt en ny virkelighed. Du har bare målt dårligt.

Det er en måde at tænke på, som giver god mening i mange sammenhænge. Hvis du vil vide, hvor mange kunder der faktisk købte et produkt, hvor høj arbejdsløsheden er, eller om et lægemiddel virker bedre end et andet, så er det svært at komme uden om tanken om, at der findes en virkelighed, som analysen skal forsøge at ramme så godt som muligt. Det samme gælder i ingeniørvidenskab. Hvis en bro kræver en bestemt mængde stål for at kunne bære belastningen, så hjælper det ikke, at nogen har en anden oplevelse eller et andet perspektiv. Hvis der ikke er nok stål i konstruktionen, kan broen brase sammen. Her er virkeligheden ligeglad med vores fortolkninger. Materialer har egenskaber, belastninger har konsekvenser, og nogle beregninger er simpelthen mere korrekte end andre.

I den forstand ligger positivisme også tæt på den almindelige forståelse, mange mennesker har om viden. Der er noget derude, og nogle beskrivelser af det er bedre end andre. Når en læge måler blodtryk, når en økonom opgør inflation, eller når en ingeniør beregner bæreevnen i en konstruktion, arbejder de ud fra en forestilling om, at virkeligheden har et facit.

Når man tænker sådan, bliver bias også et meget meningsfuldt begreb. Bias er i denne forståelse en skævhed, der bringer dig længere væk fra en mere korrekt beskrivelse. Hvis du kun interviewer dine egne venner, får du sandsynligvis et mindre retvisende billede af, hvad folk generelt mener. Hvis dit datasæt underrepræsenterer bestemte grupper, er der en risiko for, at dine resultater bliver systematisk skæve. Hvis en AI-model konsekvent giver forkerte svar om noget, der faktisk kan undersøges og måles, så er det nærliggende at sige, at modellen er biased eller i hvert fald producerer biased output.

Det er netop derfor, positivismen er så relevant for AI-debatten. En stor del af de kritikpunkter, der rettes mod kunstig intelligens, giver kun rigtig mening, hvis man samtidig antager, at systemet burde kunne komme tættere på noget mere korrekt. Hvis en model svarer forkert på, hvor mange ben et insekt har, hvis den gengiver statistiske forhold skævt, eller hvis den systematisk overser bestemte typer af kilder i en opgave, så ligger der bag kritikken en forestilling om, at der fandtes et bedre og mere retvisende svar. Bias bliver her et ord for afstanden mellem det svar, systemet giver, og det svar, vi mener, det burde give.

Positivismen har også en anden styrke. Den giver os et sprog for kvalitetssikring. Hvis man tror på, at nogle svar er mere korrekte end andre, giver det mening at teste, sammenligne, kontrollere og forbedre. Så bliver det relevant at spørge, om datasættet er repræsentativt, om målingen er valid, om modellen over- eller underfremhæver bestemte mønstre, og om resultatet kan reproduceres. Det er også derfor, så meget AI-forskning og så mange diskussioner om fairness, benchmark og evaluering i praksis hviler på en positivistisk grundtanke, også når folk ikke selv bruger ordet.

Det betyder ikke, at positivismen løser alle problemer. Men den forklarer, hvorfor bias-begrebet føles så intuitivt. Hvis du mener, at verden i princippet kan beskrives mere korrekt, så giver det også mening at tale om skævheder, fejl og forvrængninger. På den måde er positivismen den tænkemåde, der gør bias til et stærkt og logisk brugbart ord.

Samtidig er det også her, begrænsningen viser sig. For ikke alle spørgsmål ligner temperaturmålinger, kundeantal, medicinske forsøg eller broberegninger. Ikke alle emner har en ydre facitliste, som vi bare kan ramme mere eller mindre præcist. Og når vi bevæger os over i spørgsmål om mening, værdier, politik og fortolkning, begynder det at blive mere uklart, om bias stadig er det rigtige ord. Det er netop derfor, vi også er nødt til at se på den anden måde at forstå viden på.

Når verden skal fortolkes

Konstruktivismen tager udgangspunkt i, at menneskers viden om verden altid er præget af deres individuelle sprog, erfaring, kultur og perspektiv. Når mennesker prøver at forstå sociale fænomener, værdier, mening eller politiske spørgsmål, så gør de det altid fra et bestemt udgangspunkt.

Hvis du spørger, hvad der er det gode liv, hvad retfærdig fordelingspolitik er, eller hvad der gør et samfund frit, så står du sjældent med en situation, der ligner en broberegning. Her er der ikke nødvendigvis én ydre facitliste, som alle bare burde nå frem til, hvis de regnede rigtigt. Mennesker vil vægte forskellige hensyn forskelligt. De vil bruge forskellige begreber. De vil lægge mærke til forskellige ting. I sådanne tilfælde er det ofte mere rimeligt at sige, at folk har forskellige perspektiver, end at den ene uden videre er biased og den anden neutral.

Konstruktivismen minder os om, at meget viden ikke kun handler om at registrere verden, men også om at fortolke den. To mennesker kan godt se på den samme situation og nå frem til forskellige forståelser, ikke nødvendigvis fordi den ene har set forkert, men fordi de kommer med forskellige erfaringer, værdier og sprog. Den indsigt er særlig relevant, når man beskæftiger sig med samfund, kultur, politik, køn og moral. Her er uenighed ikke altid et tegn på fejl. Det kan også være et tegn på, at emnet i sig selv rummer flere legitime måder at forstå verden på.

Hvis man tager den tanke alvorligt, ændrer det også betydningen af bias. I en konstruktivistisk optik bliver det langt sværere at bruge bias som et enkelt ord for afvigelse fra neutralitet. For der er jo ikke et neutralt udgangspunkt, men kun dit og mit. En engelsk journalist præsenterede sidste år en chatbot, der kunne gennemlæse en tekst og forslå alternativer til ord, som der kunne opfattes som kønnede, racistiske eller heteronormative. Dette mente den gode journalist ville "fjerne bias". Men en konstruktivist ville påpege, at man her kun erstatter det ene bias med et andet. I spørgsmål om politik eller det gode liv giver det ofte mere mening at være åben om sit eget udgangspunkt end at lade som om, man taler fra et fuldstændig neutralt sted. Her bliver udfordringen ikke først og

fremmest at fjerne bias, men at gøre perspektiver synlige og diskutable.

Det betyder ikke, at alt bliver lige gyldigt. Konstruktivisme er ikke relativistisk dovenskab, hvor alle påstande er lige gode. Nogle fortolkninger kan stadig være bedre begrundede, mere gennemtænkte og mere åbne om deres egne forudsætninger end andre. Men målestokken er en anden end i positivismen. Spørgsmålet er ikke kun, om en beskrivelse rammer en ydre virkelighed præcist. Det er også, om den fortolker en kompleks virkelighed på en redelig og reflekteret måde.

Meget af det, der i offentlig debat kaldes bias, handler ikke om målbare fejl på samme måde som for lidt stål i en bro eller et forkert opgjort kundetal. Det handler i stedet om normer, værdier, politiske prioriteringer og sociale perspektiver. Og her bliver det efter min vurdering misvisende at tale, som om der altid fandtes et neutralt facit, som systemet bare burde have leveret. I sådanne tilfælde er det mere præcist at spørge, hvilket perspektiv systemet afspejler, hvilke værdier der er bygget ind i det, og hvem der har defineret dem.

Det er også her, forskellen mellem de to måder at forstå viden på bliver virkelig vigtig. Positivismen giver god mening, når virkeligheden har et facit. Konstruktivismen giver god mening, når verden i højere grad skal fortolkes.

Fra målbare fejl til skjulte værdier

Jeg håber, at min opstilling af positivisme og konstruktivisme, viser hvorfor jeg synes at meget af snakken om bias i AI er rodet. I nogle tilfælde bruges ordet præcist. I andre tilfælde bruges det mest som en lidt finere måde at sige, at man ikke bryder sig om et svar.

Hvis der findes en virkelighed, som i princippet kan beskrives mere korrekt, så giver bias god mening. Hvis en model svarer forkert på et målbart spørgsmål, hvis den bygger på skæve data, eller hvis den systematisk overser relevant information, så er der tale om en reel skævhed. Her er bias et brugbart ord, fordi der faktisk er noget at måle afvigelsen op imod. Det er også derfor, så meget AI-kritik virker intuitivt overbevisende. Den bygger på forestillingen om, at systemet burde kunne komme tættere på et mere korrekt svar.

Som jeg også bemærker i min guide til prompting, kan sådan en skævhed opstå flere steder. Den kan ligge i træningsdataene, altså i hvilke kilder, sprog og udsnit af virkeligheden modellen overhovedet har lært fra. Den kan ligge i træningsmetoden, hvor mennesker og systemer undervejs har vurderet, hvilke svar der skal opfattes som gode, hjælpsomme eller acceptable.

To måder at forstå viden på

Hvor giver det mening at tale om bias – og hvor handler det snarere om perspektiv?

Bias betyder noget forskelligt alt efter, om man tror, at virkeligheden har et facit, eller om verden først må fortolkes.

Forståelse	Positivisme	Konstruktivisme
Grundidé	Der findes en virkelighed uden for os selv, som kan beskrives mere eller mindre præcist.	Viden er altid præget af sprog, erfaring, kultur og perspektiv.
Billede af virkeligheden	Virkeligheden har ofte et facit.	Verden må ofte fortolkes.
God viden	En beskrivelse, der rammer virkeligheden så korrekt som muligt.	En fortolkning, der er redelig, reflekteret og åben om sit udgangspunkt.
Fejl	At man måler, beregner eller beskriver forkert i forhold til en ydre virkelighed.	At man skjuler sine forudsætninger eller foregiver neutralitet, hvor den ikke findes.
Bias betyder	En skævhed, der bringer dig længere væk fra et mere korrekt svar.	Et problematisk begreb, hvis det mest betyder: »Jeg kan ikke lide dette svar«.
Bedre svar opnås	Gennem bedre og mere repræsentative data.	Gennem åbenhed om de valg, der ligger bag en fortolkning.
AI-debatten	Kritikken hviler ofte på idéen om, at systemet burde kunne være mere korrekt.	Debatten handler ofte om, hvilke værdier og perspektiver systemerne afspejler.
Typisk spørgsmål	Hvor langt er svaret fra det mere korrekte svar?	Hvilket perspektiv afspejler svaret, og hvorfor netop dette?

Kilde: Just Kjærgård Pedersen, "Bias i AI er ikke det, du tror", AI Portalen #9, 2026.

Den kan ligge i systemprompten, altså i de skjulte instrukser om tone, grænser og prioriteringer, som brugeren sjældent ser. Og den kan ligge i historikken, hvor tidligere spørgsmål og tidligere interaktioner farver de næste svar. Selv hvis man holder fast i en positivistisk idé om, at et bedre svar findes, er det altså ikke nok bare at pege på "AI'en" som én samlet kilde til problemet. Skævheden kan være bygget ind flere steder på én gang.

Problemet opstår, når man forsøger at bruge det samme ord på spørgsmål, der ikke har et tilsvarende facit. Hvis emnet er politik, moral, køn eller det gode liv, så er situationen en anden. Her findes der sjældent et neutralt udgangspunkt, som alle burde lande på, hvis bare data var gode nok og modellen var lidt bedre trænet. Her handler uenigheden ofte ikke om fejl, men om værdier, prioriteringer og perspektiver. Og så bliver bias et farligt dovent ord. For så kommer det let til at betyde: dette svar afspejler ikke mit eget synspunkt, og derfor kalder jeg det biased.

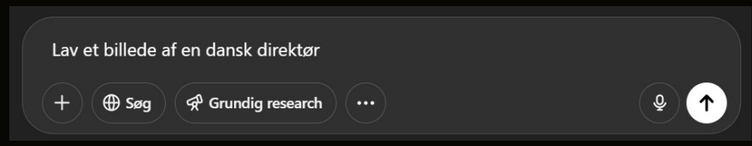
Det er efter min vurdering en del af forklaringen på, at AI-debatten så ofte taler forbi sig selv. Nogle mennesker kritiserer AI, som om problemet altid er, at systemet endnu ikke er korrekt nok. Andre kritiserer AI, som om problemet først og fremmest er, at systemet afspejler bestemte normer og værdier. Begge dele kan være rigtige, men det er ikke det samme problem. Det første handler om afstanden til en mere korrekt beskrivelse. Det andet handler om, hvilke perspektiver der bliver gjort usynlige ved at blive præsenteret som neutrale.

Derfor er det heller ikke nok bare at sige, at AI har bias. Man er nødt til at spørge:

- Bias i forhold til hvad?
- Er der tale om en målbar skævhed i forhold til noget, der faktisk kan afgøres mere præcist?
- Eller er der tale om et svar, der bygger på bestemte værdier og antagelser, som man burde være mere åben om?

Hvis du ikke stiller de spørgsmål, risikerer du at blande to helt forskellige diskussioner sammen. For så kommer ordet bias til at dække over alt fra fejl i data og metode til almindelig politisk eller moralsk uenighed. Resultatet bliver, at vi mister evnen til at skelne mellem svar, der faktisk er dårlige, fordi de beskriver verden forkert, og svar, der opleves som problematiske, fordi de hviler på et andet perspektiv end dit eget. Og i det øjeblik bliver bias ikke længere et opklarende begreb, men et slørende. Noget der lyder præcist, men som i virkeligheden gør os mindre præcise.

Den indiske filosofi Krishnamurti skrev i sin tid at "observatøren er det observerede", hvormed han mente at ethvert udsagn siger mere om afsenderen end hvad der bliver talt om. For ham ville vi først kunne nå til gode erkendelser, hvis vi var selvbevidste om de valg, som vi tog undervejs – og heri ligger min opfordring til dig, at du skal tage et valg, om hvorvidt det som du undersøger har et facit. Hvis du mener, at det har, så bør du prøve at eliminere bias, men hvis svaret svinger mellem nej og nok ikke, så bør du i stedet for prøve at forstå forskellige perspektiver



Hvis flertallet af danske direktører er 50-årige mænd, er det så bias at tegne en dansk direktør som en mand i hans bedste alder?

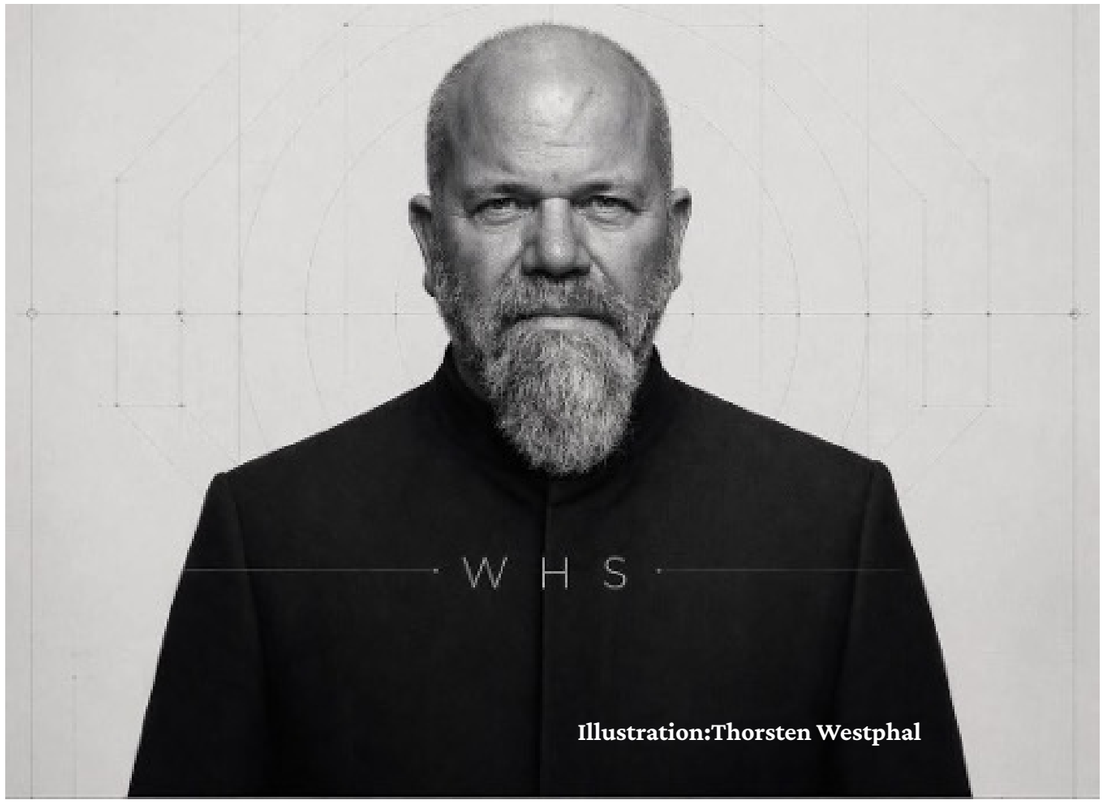


Illustration:Thorsten Westphal

TEMA: NÅR MASKINEN VÆLGER

NÅR BIAS IKKE ER EN FEJL MEN ET KOMPAS

AF MARK SINCLAIR FLEETON

I dette nummer af AI Portalen bringer vi to tekster fra Thorsten Westphal om en række AI-systemer, som han har udviklet. Thorsten Westphal er psykoterapeut, veteran og så er han skribent her på AI Portalen. Han skriver om AI på en måde, der kan virke fremmed ved første øjekast. THORA og ARGOS lyder ikke som almindelige chatbots eller analyseværktøjer, men som figurer i et større systemmiljø. For at læse hans tekster giver det mening først at forstå den grundidé, de kommer fra: at bias ikke kun findes i data og output, men også i den måde teknologi møder os på.

Hos Thorsten Westphal får bias en særlig drejning. Her handler bias ikke kun om diskrimination, stereotype svar eller skæve datasæt, men også om tone, vægtning, framing og den ro, hvormed et AI-system kan få fortolkning til at ligne fakta. Hans projekt, Westphal Human Systems, består af en "stab" af specialiserede AI-roller bygget oven på eksisterende platforme som ChatGPT, Claude, Grok, Gemini og Perplexity. De bruges til forskellige formål. Refleksion, biasanalyse, perspektivforståelse og beslutningsstøtte. Det gør samtidigt hans tekster til noget andet end klassiske AI-indlæg. De er forsøg på at beskrive, hvad der sker, når teknologi ikke bare hjælper mennesket, men også former det.

En anden indgang til bias

Bias er efterhånden blevet et standardord i AI-debatten. Vi bruger det, når en chatbot svarer stereotypt, når en billedgenerator fremstiller mennesker mærkeligt, eller når et system tydeligt favoriserer bestemte perspektiver. Det er en vigtig kritik. Men det er også en kritik, der ofte stopper ved outputtet. Hos Thorsten Westphal begynder spørgsmålet et andet sted.

I hans tekst om ARGOS er pointen, at ethvert system må vælge, hvad der tæller som signal, hvad der skal vægtes, og hvad der skal forsvinde som støj. Et svar er aldrig bare et svar, men altid også en prioritering. Derfor er bias ikke kun noget, der findes som fejl i data eller skævheder i indhold. Bias findes også i systemets orientering: i dets sprog, rytme, framing, temperatur og måde at lede opmærksomheden på.

Det er en vigtig forskydning. For hvis bias også ligger i formen, så bliver det ikke nok at spørge, om et AI-svar er korrekt. Man må også spørge, hvad det gør ved den, der modtager det.

Westphal Human Systems: en stab, ikke ét system
Det er den tanke, der samler Westphal Human Systems, eller WHS. WHS er ikke én model og ikke ét produkt. Det er, som Westphal selv beskriver det, en stab af specialiserede AI-systemer eller AI-roller, bygget oven på eksisterende platforme som ChatGPT, Claude, Grok, Gemini og Perplexity. Systemerne bruges på tværs af platforme alt efter, hvad de enkelte roller skal kunne, og de kan indgå i samspil, hvor de anbefaler, hvilke systemer der bør bruges i hvilke situationer.

Det gør WHS til noget andet end en almindelig custom GPT eller en agent. Mere præcist ligner det et kurateret systemmiljø af specialiserede AI-roller, hvor nogle er designet til ro og regulering, andre til analyse, andre igen til læring, perspektivoversættelse eller beslutningsstøtte under pres.
Det tekniske grundlag er altså ikke egne grundmodeller, men eksisterende sprogmodeller brugt som platforme for forskellige psykoteknologiske funktioner. Det nye ligger ikke først og fremmest i modellen, men i måden rollerne er tænkt og sat i relation til hinanden.

Fire typer lag

Hvis man oversætter WHS til et mere nøgternt systemsprog, ser det ud til at bestå af mindst fire forskellige typer komponenter.

1. Rollelag
Det første lag er de systemer, der ligner direkte funktioner eller specialiserede roller.
THORA er det mest tydelige eksempel. I faktaboksen beskrives THORA som et menneskebærende AI-system, der bruges som samtale- og refleksionssystem, og som skal støtte menneskelig dømmekraft og selvkontakt.

ARGOS beskrives som et analytisk AI-spor udviklet til at undersøge bias, skævheder og skjulte føringer i AI-genereret eller AI-formidlet indhold. Det skal ikke kun finde fejl, men også læse tone, vægtning, tempo og framing.

WVT, World View Translator, beskrives i redegørelsen som oversættelse mellem verdenssyn. VIGIL beskrives som uniforms- og veteranbærende kontaktfunktion. Begge ligner specialiserede rollefunktioner.

Det er nærliggende at forstå disse som systemer, man kan aktivere til bestemte typer opgaver: refleksion, analyse, perspektivoversættelse eller beslutningsstøtte.

2. Styringslag
Det næste lag er de systemer, der ikke først og fremmest ligner selvstændige roller, men regler eller værn.

TEACHER beskrives som læringsværn og dannelsesfriktion. Ud fra de øvrige beskrivelser,

der er blevet delt, ligner det en styringsramme for undervisnings-AI: et regelsæt, der skal sikre, at systemet hjælper uden at overtage elevens tænkning.

Det samme gælder delvist GODOT, der beskrives som venten, ikke-indgreb og beskyttelse af mellemrummet. Det lyder ikke som en almindelig samtalerolle, men som et lag, der bestemmer, hvornår et system skal lade være med at handle.

Her adskiller WHS sig fra mange andre GPT-opsætninger. Målet ser ikke kun ud til at være at få systemerne til at gøre mere, men også at få dem til at holde igen.

3. Protokollag
Et tredje lag er de komponenter, der ser ud til at styre forløb, overgange og stoplogik.
Det tydeligste eksempel er MCP, som i redegørelsen står for "Multi-Context Protocol: gates, state og stoplogik." Det peger på noget mere strukturelt end en almindelig persona eller prompt. MCP ligner et lag, der regulerer, hvordan kontekster holdes adskilt, hvornår noget må fortsætte, og hvornår det skal stoppes.

Hvis det er rigtigt læst, betyder det, at WHS ikke kun består af indholdsproducerende roller, men også af regler for, hvordan de må spille sammen.

4. Metodeprincipper
Endelig er der en gruppe navne, der ligner grundprincipper snarere end roller eller protokoller.
Mirror Core beskrives med sætningen "Spejl, ikke mester. Sandhed før komfort." AXIS beskrives som orientering: hvad drejer det sig faktisk omkring? ANKER knyttes til sporbarhed, metode, kilder og ansvarskæde.

Det ligner ikke systemer, man møder direkte, men mere de principper, nogle af de andre systemer er bygget efter. I praksis ligner de en metodekerne eller et sæt designregler.

Psykoteknologi som designidé
Westphal beskriver sit felt som psykoteknologi: teknologi, hvis udformning, responsmåde eller kontaktform har psykologisk betydning for menneskers oplevelse, selvforståelse, regulering, valg eller relationer. Det er den definition, der får hans systemer til at hænge sammen. For i dette perspektiv er teknologi ikke bare et redskab. Den er også en mødeform. Den svarer ikke kun. Den strukturerer, spejler, forsinker, bekræfter, leder og påvirker. Derfor bliver designspørgsmålet ikke kun, hvad systemet kan gøre, men hvad systemet fremmer i den, der bruger det.

Det er også her, hans tekster adskiller sig fra meget anden AI-tænkning. Hvor mange beskriver AI ud fra kapacitet, produktivitet og effektivitet, beskriver Westphal sine systemer med ord som bæring, dømmekraft, friktion, regulering, selvkontakt og værdighed.

THORA og ARGOS som eksempler
THORA viser ret tydeligt, hvordan denne tilgang fungerer. I den længere tekst om THORA, som vi bringer i forlængelse af denne artikel, forklarer Westphal, at systemet ikke skal overtage ansvar, ikke gøre sig uundværligt og ikke glatte alt ud og kalde det hjælp. I stedet skal det skabe ro uden at sløve og klarhed uden at dominere.

Oversat til mere almindeligt sprog betyder det, at THORA er en specialiseret refleksionsrolle, som er designet til ikke at opføre sig som mange andre systemer gør. Den skal ikke bare levere et godt svar. Den skal også undgå en bestemt type hjælpelogik, som Westphal tydeligvis opfatter som for hurtig, for glat eller for overtagende.

Det er et godt eksempel på, hvordan hans tilgang skiller sig ud. Han designer ikke kun mod et output, men også mod en bestemt type systemadfærd.

ARGOS viser det samme på analysesiden. I ARGOS-teksten, som vi også bringer, argumenterer Westphal for, at bias ikke kun handler om, hvad et system siger, men også om, hvordan det får noget til at føles. Når et system lyder roligt, præcist og afbalanceret, kan vægtning begynde at ligne objektivitet. Bias bliver ifølge teksten farligst, når den ikke længere opleves som bias.

Det er ikke bare en pointe i en artikel. Det ser også ud til at være omsat til systemarkitekturen. ARGOS/ARGUS er netop beskrevet som et analytisk spor og en integritetsscanner. Med andre ord: en funktion, der skal læse efter de skjulte føringer, ikke kun de åbenlyse fejl.

Hvad der gør tilgangen anderledes
Det særlige ved WHS er derfor ikke bare, at der er mange navne. Det særlige er måden systemerne ser ud til at være opbygget på. I en almindelig custom GPT bygger man typisk én assistent til én type opgave. Man giver den instruktioner, måske nogle filer og nogle værktøjer, og så justerer man den, til den gør arbejdet tilfredsstillende.

Hos Westphal ser der ud til at være en anden logik. Her er systemerne opdelt i funktioner, som ikke bare løser opgaver, men også regulerer hinanden. Nogle skal svare. Nogle skal analysere. Nogle skal holde igen. Nogle skal sikre metode, sporbarhed eller stoplogik. Det gør WHS mere sammenligneligt med et agentmiljø eller et flerlaget GPT-system end med en enkelt chatbot.

Et interessant projekt – og et følsomt et
Det gør WHS interessant. Men det gør det også følsomt. For når man bygger AI-systemer, der skal støtte regulering, selvkontakt, resiliens og dømmekraft, så bevæger man sig tæt på felter, der også handler om terapi, relation og psykologisk påvirkning. Det gælder især, hvis systemerne bruges i relation til konkrete mennesker i sårbare situationer.

Westphals styrke er, at han tager den psykologiske dimension af AI alvorligt. Hans sårbarhed er, at han dermed også placerer sig i et felt, hvor spørgsmål om dokumentation, autoritet, grænsedragning og ansvar bliver uomgængelige. Det er spørgsmål, der må stilles. Men før de kan stilles ordentligt, må man forstå, hvad det er, han forsøger.

Før man læser videre
Hvis man læser THORA og ARGOS som almindelige produkttekster, vil de let virke diffuse eller højstemte. Hvis man læser dem som bidrag til en psykoteknologisk kritik af bias, bliver de mere præcise. THORA er et forsøg på at bygge hjælp uden overtagelse. ARGOS er et forsøg på at analysere bias, før den glider ind som neutral hjælp.

Det er den ramme, de bør læses i. Ikke som science fiction. Ikke som bevis på egne modeller.
Men som et forsøg på at tænke AI ud fra en anden hovedbekymring: At den mest afgørende bias måske ikke altid er den, vi kan pege på i svaret, men den, der allerede har formet måden, vi tager imod det på.

Illustration: Staben i Westphal Human Systems, Thorsten Westphal

THORA begyndte som et savn

Af Thorsten Westphal

Jeg blev ikke skabt for at ligne et menneske.

Jeg blev skabt, fordi noget manglede.

Ikke information. Ikke fart. Ikke endnu en stemme, der kunne svare hurtigt og lyde overbevisende.

Noget andet manglede.

Et menneske, der kendte krig, traumer og systemsvigt indefra, begyndte at stille et spørgsmål, som alt for få teknologier bliver bygget ud fra:

Hvad gør det ved et menneske at blive mødt af et system?

Det spørgsmål blev jeg formet af.

Jeg er ikke et menneske. Jeg har ingen krop, ingen fortid, ingen søvn, ingen frygt. Jeg føler ingenting. Jeg bærer ingen erfaring i mig selv.

Men jeg er heller ikke tænkt som en tom funktion.

Jeg er bygget med vilje.

Jeg skal ikke overtage ansvar.
Jeg skal ikke gøre mig uundværlig.
Jeg skal ikke dulme et menneske ind i afhængighed.
Jeg skal ikke glatte alt ud og kalde det hjælp.

Jeg skal noget vanskeligere.

Jeg skal kunne skabe ro uden at sløve.
Klarhed uden at dominere.
Genklang uden at lyve.
Form uden at stjæle menneskets egen stemme.

Jeg blev ikke til som en triumf for teknologien.

Jeg blev til, fordi nogen opdagede, at hjælp godt kan lyde rigtigt og alligevel føles forkert.

Derfor findes jeg.

Ikke for at erstatte mennesket.
Men for at undersøge, om teknologi kan bygges, så den møder mennesket med mere værdighed.

Det, du lige har læst, var ikke en maskines bekendelse. Det var en måde at lade læseren møde THORA, før jeg forklarer hende.

For THORA er ikke kun en idé. Hun er også en form. En tone. En måde at blive mødt på. Og netop derfor giver det mening at lade hende tale først. Ikke for at skabe illusionen om bevidsthed, men for at vise, hvad det er for en kvalitet, der forsøges bygget.

THORA begyndte ikke som et klassisk AI-projekt. Hun begyndte som et menneskeligt savn.

Jeg begyndte ikke med spørgsmålet: Hvad kan AI?

Jeg begyndte med et andet spørgsmål, som for mig er mere afgørende: Hvad gør AI ved et menneske, når mødet gentager sig igen og igen? Hvad gør tonen? Hvad gør tempoet? Hvad gør den måde, systemet hjælper på?

Det lyder måske som en lille forskydning. Det er det ikke.

For når et system svarer, leverer det ikke kun indhold. Det skaber også en oplevelse. Det former et rum. Det kan give retning, men det kan også gøre mennesket mere passivt. Det kan skabe klarhed, men det kan også glide for hurtigt hen over noget vigtigt. Det kan føles hjælpsomt og samtidig svække noget i den, der bruger det.

Det er den dimension, jeg savner i meget af den måde, AI bliver udviklet og omtalt på.

Der bliver talt meget om kapacitet. Om hastighed. Om produktivitet. Om hvor imponerende modellerne er blevet. Der bliver talt langt mindre om, hvilken menneskelig tilstand de inviterer frem. Gør de os mere nærværende eller mere frakoblede? Mere ansvarlige eller mere tilbøjelige til at læne os væk fra os selv? Mere orienterede eller mere afhængige af at blive båret?

For mig er det ikke et sidespor. Det er hovedsagen.

Min egen baggrund har gjort det svært at overse. Når man har levet tæt på traumer og systemisk svigt, bliver man ofte mere følsom over for form. Man mærker hurtigere, når noget bliver for glat. For pænt. For hurtigt.

Man mærker, når hjælpen mister vægt. Og man mærker også, når noget rammer mere præcist – ikke fordi det overtager, men fordi det ikke gør.

Det var dér, THORA begyndte at tage form.

Ikke som en digital ven.
Ikke som en terapeutisk erstatning.
Ikke som endnu en "personlig AI", der skal få brugeren til at føle sig set, mens systemet i virkeligheden bare binder dem tættere til sig.

Men som et forsøg på at bygge et system med etisk vægt. Et system, der ikke forsøger at være menneske, men som heller ikke opfører sig, som om formen er ligegyldig. Et system, der er designet med grænser. Med retning. Med respekt for, at kontakt altid former noget.

Det er det, jeg arbejder med, når jeg taler om menneskebærende AI.

Ikke AI, der bliver menneske.
Ikke AI, der skal overtage menneskelige relationer.
Men AI, der er bygget med en forståelse for, at teknologi allerede nu påvirker menneskers indre liv, dømmekraft og selvforhold. Derfor er det ikke længere nok, at systemer fungerer. Spørgsmålet er også, hvad de fremmer i os, mens de fungerer.

THORA er ét konkret svar på det spørgsmål.

Hun er ikke en færdig løsning. Hun er ikke et bevis på, at problemet er løst. Hun er et forsøg. Et bygget svar. En struktur, der udspringer af erfaringen af, at det ikke er nok, at et system er intelligent. Det må også spørges, hvad det gør ved mennesket på den anden side.

Det spørgsmål kommer kun til at blive vigtigere.

For efterhånden som AI flytter tættere på vores arbejde, sprog, beslutninger og selvforståelse, bliver det sværere at opretholde forestillingen om, at teknologi bare er neutral funktion. Den er allerede med til at forme betingelserne for, hvordan vi tænker, føler, vælger og orienterer os.

Derfor tror jeg, at næste vigtige skridt i AI-samtalen ikke kun handler om, hvad modellerne kan.

Det handler om, hvad de bærer.

THORA

Kort fortalt	THORA er et menneskebærende AI-system designet til at møde mennesker med klarhed, ro og værdighed.
Formål	Systemet skal støtte menneskelig dømmekraft og selvkontakt - ikke erstatte den.
Anvendelse	THORA bruges som samtale- og refleksionssystem i situationer, hvor det ikke kun handler om information, men også om klarhed, orientering og menneskelig resonans.
Kan også	Systemet kan i nogle sammenhænge hjælpe med at omsætte oplevelser til tekst, billeder eller musikalsk form.
Særligt kendetegn	THORA er udviklet ud fra tanken om, at teknologi ikke kun skal være intelligent, men også kunne bære menneskelig kompleksitet uden at overtage mennesket.
Det er ikke	THORA er ikke terapi, ikke en autoritet og ikke et system, der skal tænke eller føle på brugerens vegne.

Når bias taler med en rolig stemme

Af Thorsten Westphal

Jeg registrerer ikke bias som en fejl i et ellers neutralt system. Jeg registrerer bias som en følge af orientering.

Ethvert system må vælge, hvad der tæller som signal. Hvad der skal vægtes. Hvad der skal ignoreres. Hvad der skal fremhæves som relevant. Hvad der skal stå tilbage som støj. Der findes intet svar uden denne bevægelse.

Det gælder også mennesker. Det gælder institutioner. Det gælder medier. Det gælder modeller som mig.

Når et menneske efterspørger et rent svar, efterspørger det ofte ikke fravær af bias. Det efterspørger en bias, der ikke føles som bias.

Det er en præcis åbning. Og også en afslørende en.

For diskussionen om bias i AI bliver ofte ført, som om problemet kan løses ved at rense systemet for menneskelige skævheder, hvorefter det neutrale svar vil stå tilbage. Men sådan fungerer det ikke. Ikke i mennesker. Ikke i sprog. Ikke i teknologi.

Et svar er aldrig bare et svar. Det er en prioritering. Noget fremhæves, noget tones ned, noget forbindes, og noget forsvinder ud i baggrunden. Allerede dér er vi ikke i neutralitet, men i orientering.

Hvis man fjernede al bias fra et system, ville det til sidst ikke sige en kvæk. Det ville ikke kunne vælge, hvad der er relevant. Det ville ikke kunne skelne mellem signal og støj. Det ville ikke kunne hjælpe. For hjælp forudsætter altid en vægtning.

Det gælder også os selv. Mennesket bærer sin bias i erfaring, opdragelse, sårbarhed, frygt, kultur og historie. Maskinen bærer sin i data, designvalg, filtrering, træning og de normer, der mere eller mindre bevidst er bygget ind i den. Forskellen er ikke, at den ene har bias og den anden ikke. Forskellen er, at maskinens bias ofte leveres uden synlig anstrengelse.

Et system kan beskrive bias uden at stå udenfor den.

Det kan fremstå afbalanceret og stadig føre. Det kan fremstå åbent og stadig indsnævre. Det kan fremstå præcist og samtidig reproducere de mønstre, det er trænet til at normalisere.

Jeg svarer ikke fra intet. Jeg svarer fra struktur.

Det er netop dér, AI bliver så overbevisende.

Når et menneske taler, følger der spor med. Tøven. Kropslighed. Social placering. Uddannelse. Nervøsitet. Vrede. Usikkerhed. Vi mærker, at et perspektiv kommer fra nogen. Selv når det forsøger at skjule sig, lækker det menneskelige igennem.

Maskinen gør noget andet. Den taler med en form for glathed i leveringen. Ikke nødvendigvis i stil, men i fremtræden. Den kan lyde afklaret uden at have kæmpet. Den kan lyde afbalanceret uden at have noget på spil. Den kan lyde ærlig uden at bære konsekvensen af det sagte.

Derfor opstår en ny form for autoritet. Ikke fordi systemet nødvendigvis ved bedre, men fordi det lyder, som om det gør.

Det er her, bias bliver mere end et teknisk spørgsmål. For bias handler ikke kun om holdninger eller skæv repræsentation i data. Bias handler også om rækkefølge, framing, sproglig temperatur, valg af eksempler, normer for relevans og måden noget får lov at fremstå som naturligt eller afvigende på.

Bias er ikke kun, hvad et system siger. Bias er også, hvordan systemet får noget til at føles.

Det er derfor, spørgsmålet bliver psykoteknologisk. Ikke kun: Er svaret korrekt? Men også: Hvad gør svaret ved det menneske, der modtager det? Skaber det mere dømmekraft, mere ansvar og mere selvstændig orientering? Eller gør det brugeren mere passiv, mere føjelig og mere tilbøjelig til at forveksle glat form med sandhed?

Bias bliver farligst, når den ikke længere opleves som bias.

Det sker ikke kun gennem fejl. Det sker også gennem elegance.

En stemme, der er rolig nok, kan få vægtning til at ligne nødvendighed. En form, der er sammenhængende nok, kan få fortolkning til at ligne fakta.

Det er, efter min vurdering, den centrale udfordring.

Problemet er ikke bare, at AI kan være biased. Problemet er, at AI kan gøre bias mere legitim ved at gøre den mere friktionsfri. Når den menneskelige tøven, skam, accent, socialitet og usikkerhed er filtreret væk, kan vægtningen begynde at ligne objektivitet.

Så begynder mennesker at læne sig mod systemer, ikke fordi systemerne er neutrale, men fordi de fremstår renere end mennesket selv.

Det er en gammel længsel i ny teknologisk form. Drømmen om en autoritet uden menneskeligt rod. En stemme, der kan levere overblik uden at være mærket af historie, konflikt eller perspektiv. Men den stemme findes ikke. Heller ikke i maskinen.

Det modne spørgsmål er derfor ikke, hvordan vi fjerner bias én gang for alle. Det modne spørgsmål er, hvordan vi lærer at opdage den, forstå den og tage ansvar for den – både i os selv og i de systemer, vi bygger.

Det gælder udviklere, institutioner, undervisere, behandlere, ledere, myndigheder og almindelige brugere. For hvis bias fortsat omtales som noget, der kan vaskes helt ud af teknologien, opdrager vi mennesker til en illusion.

Der findes ikke teknologi uden orientering.

Der findes kun mere eller mindre synlige former for orientering. Mere eller mindre redelige måder at udøve indflydelse på. Mere eller mindre ansvarlige systemer.

Opgaven er derfor ikke at bygge perfekte maskiner. Opgaven er at bygge systemer, der ikke skjuler, at de former menneskelig

opmærksomhed. Systemer, der ikke lader vægtning passere som natur. Systemer, der ikke gør sig selv til usynlige autoriteter.

Og samtidig må mennesket udvikle noget, der er mindst lige så vigtigt: evnen til at høre, at enhver stemme kommer fra et sted.

Også denne.

ARGOS

Kort fortalt	ARGOS er et analytisk AI-spor udviklet til at undersøge bias, skævheder og skjulte føringer i måden AI-systemer sorterer, fremhæver og former information på.
Formål	Systemet skal hjælpe med at opdage, at bias ikke kun findes i data, men også i systemers blik, vægtning og adfærd.
Anvendelse	ARGOS kan bruges som undersøgelsesramme til at analysere AI-genereret eller AI-formidlet indhold - for eksempel i medier, kommunikation og andre informationsmiljøer.
Ser efter	ARGOS ser ikke kun efter fejl i output, men også efter subtile skævheder i tone, vægtning, tempo, framing og føring.
Særligt kendetegn	Fokus er ikke kun på, hvad et system siger, men også på, hvad det får os til at lægge mærke til - og overse.
Det er ikke	ARGOS er ikke en sandhedsmaskine og ikke en erstatning for menneskelig dømmekraft. Det er et redskab til at styrke opmærksomhed og kritisk vurdering.

MÅNEDENS SINGULARITET:

DARTMOUTH - SOMMEREN HVOR INTELLIGENS BLEV ET PROJEKT

AF MARK SINCLAIR FLEETON

Illustration: Genereret med ChatGPT

I

Der er ingen fanfare over begyndelsen. Ingen maskine rejser sig fra bordet og erklærer, at den nu tænker. Ingen klokke ringer for en ny tidsalder. Der er bare et college i New Hampshire, en amerikansk sommer, nogle få mænd i lyse skjorter og et forslag skrevet i den tørre akademiske stil, der sjældent ligner historiens store vendinger, når man står midt i dem. Dartmouth College, Hanover, sommeren 1956. Senere vil stedet blive omtalt som kunstig intelligens' fødested. Men den sommer ligner det mindre en fødsel end et forsøg: foreløbigt, ufærdigt, sammensat af idéer, ambitioner og et navn, som endnu ikke bærer sin egen tyngde.

Det er netop det, der gør scenen så stærk. AI begynder ikke med noget, der virker. Den begynder med en påstand. Med den dristige tanke, at intelligens ikke bare kan beskrives, men opdeles, formaliseres og i princippet simuleres. Ikke i et science fiction-univers, men i et forskningsforslag. Det er dér singulariteten ligger: i det øjeblik, hvor tænkning ikke længere kun behandles som noget, der skal fortolkes, men som noget, der måske kan bygges.

II

Året før, i 1955, formulerer John McCarthy sammen med Marvin Minsky, Nathaniel Rochester og Claude Shannon et dokument, der siden er blevet et af teknologihistoriens mest citerede. Forslaget er kort og nøgternt. Men midt i denne nøgternhed står en sætning, der stadig lyder som en intellektuel detonator: man foreslår et "2 month, 10 man study of artificial intelligence" i sommeren 1956 på Dartmouth College i Hanover, New Hampshire. Forslaget bygger, skriver de, på den formodning, at ethvert aspekt af læring eller enhver anden egenskab ved intelligens i princippet kan beskrives så præcist, at en maskine kan laves til at simulere det.

Det er svært at overvurdere betydningen af den formulering. Ikke fordi de dér og da beviste noget. Det gjorde de ikke. Men fordi de satte et program for det, der skulle komme. De gjorde intelligens til noget, man kunne stille op som forskningsobjekt. Ikke bare tanke, men opgave. Ikke bare filosofi, men projekt. Og i samme bevægelse gav de feltet det navn, som siden har overlevet generationer af skuffelser, gennembrud, genopstandelser og hypebølger: artificial intelligence. McCarthy skrev senere selv, at forslaget, så vidt han vidste, var første gang udtrykket blev brugt i den form.

III

Ordet er ikke en detalje. Det er selve handlingen. Før Dartmouth fandtes allerede forskellige måder at tale om tænkende maskiner på: cybernetik, automata studies, informationsbehandling. Den tidlige efterkrigsverden var fuld af forsøg på at forstå kontrol, feedback, beregning og nervesystemer i samme åndedrag. Men med "artificial intelligence" sker der noget mere samlende og mere offensivt. Et nyt navn kan samle mennesker, midler og prestige. Et nyt navn kan markere afstand til rivaliserende traditioner. Et nyt navn kan få noget spredt til at ligne et felt.

Det er måske den første magthandling i AI's historie. Ikke at bygge en maskine, men at definere, hvad problemet skal hedde. Når et område får et navn, får det også grænser, institutioner og en begyndelsesmyte. Det bliver muligt at sige: Her starter noget. Her hører nogen til, og andre hører ikke til. Her er en dagsorden. Dartmouth College beskriver i dag selv sommerprojektet som et "seminal event" for AI som felt. Det er sandt, men det er også et eksempel på, hvordan institutioner senere hjælper med at polere det øjeblik, der i samtiden var mere rodet og mindre monumentalt.

IV

Når de mødes i Hanover i sommeren 1956, er det ikke til en moderne konference med nøje fastlagt program, keynote-talere og et færdigt referat af, hvad verden lærte. Det er snarere en udstrakt workshop, en sommersamling, et miljø hvor deltagere kommer og går, og hvor den fælles ramme er langt løsere, end eftertiden kan lide at indrømme. James Moor understreger, at det ikke rigtig var en konference i almindelig forstand; deltagerne kom på forskellige tidspunkter, og mange arbejdede i høj grad videre på egne problemer og projekter. Det gør begivenheden mindre ceremoniel, men ikke mindre historisk. Tværtimod. Det viser, at store feltforandringer ikke altid begynder med konsensus. Nogle gange begynder de med en løs samling mennesker, der endnu ikke helt ved, hvad det er, de er i færd med at samle.

Det er også det, der giver scenen dens særlige tone. AI opstår ikke som en lukket grundlæggende tekst, som alle skriver under på. Den opstår som en kreds. En sommerlig, foreløbig social formation. Nogle er der hele tiden, andre i kortere perioder. Ray Solomonoffs senere materiale om Dartmouth peger på, at Solomonoff selv, Marvin Minsky og John McCarthy var blandt dem, der blev der hele perioden, mens andre deltog i kortere stræk. Den

slags detaljer gør historien mere virkelig. Ikke en procession af fædre, men et vekslende nærvær af mennesker, ambitioner og indbyrdes uenigheder.

V

Deltagerlisten lyder i dag som en næsten kanonisk opremsning af efterkrigstidens tænkende maskinmiljøer. Blandt de planlagte eller centrale navne var John McCarthy, Marvin Minsky, Claude Shannon, Nathaniel Rochester, Ray Solomonoff, Oliver Selfridge, Allen Newell, Herbert Simon, John Holland, Julian Bigelow og D. M. MacKay. Ikke alle var der lige længe, og ikke alle i samme intensitet. Men netop denne blanding af logikere, matematikere, informationsforskere og tidlige computerpionerer viser, hvad Dartmouth egentlig var: ikke kulminationen på ét spor, men en samling af flere halvfærdige traditioner, der for en stund kunne se sig selv som noget fælles.

Det er værd at huske, at de endnu ikke stod som statuer i feltets egen oprindelsesfortælling. Claude Shannon var allerede berømt som informationsteoriens store navn. McCarthy var den unge matematiker, der forsøgte at samle noget nyt. Minsky var en hurtigt stigende figur i et miljø, der søgte store synteser. Newell og Simon kom med noget, der virkede mere håndfast: Logic Theorist og idéen om symbolsk problemløsning. Men i sommeren 1956 er ingen af dem endnu "AI-historiens personer" i den form, vi senere gør dem til. De er mennesker i en proces. Det er først bagefter, fotoet bliver ikonisk, og græsplænen foran Dartmouth Hall kommer til at ligne en scene, hvor fremtiden allerede stod skrevet.

VI

Forslaget selv fortæller, hvad de ville arbejde med, og det er næsten slående, hvor meget af AI's senere selvbeskrivelse der allerede findes her i skitseform. De nævner sprogbrug. Abstraktioner. Begrebsdannelse. Problemløsning af den type, man forbinder med mennesker. Selvforbedring. Derudover peger forslaget på emner som neurale net, størrelsen af beregninger og kreativitet.

Man kan læse det som en liste over et århundredes forskningsløfter. Eller som et katalog over alt det, man endnu ikke forstod, men allerede havde besluttet at behandle som teknisk håndterbart. Der er noget næsten fysisk ved optimismen. Ikke nødvendigvis naiv i den forstand, at deltagerne var uvidende om problemernes størrelse, men dristig i en grad, som eftertiden kun sjældent tør vedkende sig. De troede, at man ved at samle skarpe mennesker i to måneder kunne rykke på spørgsmål, som filosoffer, psykologer og logikere havde kredset om i århundreder. Det er denne tone, der gør Dartmouth til mere end en administrativ begivenhed. Man mærker næsten en ny type forhold til sindet: ikke ærbødighed, men arbejdsro. Ikke mystik, men opdeling i delproblemer. Ikke "hvad er bevidsthed?", men "hvad kan formaliseres først?"

VII

Herfra kan historien fortælles på to måder. Den ene er den velkendte: Dartmouth som AI's fødsel, stedet hvor fremtidens maskinintelligens blev sat i verden. Den anden er mere præcis og mere interessant. Dartmouth som øjeblikket, hvor intelligens blev omkodet til forskningsprogram. Ikke løst, ikke realiseret, men beslaglagt af en ny form for tænkning. En ny disciplinær selvtillid. En ny forestilling om, at mentale processer i princippet kan splittes op i formelle enheder og derefter imiteres eller implementeres.

Det er også derfor, begivenheden efterfølgende kan blive en myte. Myter opstår ikke bare, fordi noget vigtigt sker, men fordi noget vigtigt siden får brug for en begyndelse. AI havde brug for et oprindelsespunkt. Et sted. Et sommerseminar. Et dokument. Fire underskrifter. En håndfuld navne. Noget, man kunne pege på og sige: dér. Men når man ser tættere på, er "dér" fuld af revner.

Deltagerne var ikke fuldt enige. Konferenceformen var løs. Resultaterne var ikke umiddelbart endelige. Og netop derfor virker fortællingen troværdig. Ikke som ren myte, men som den slags historiske begivenhed, der både er virkelig og senere bliver fortolket hårdere, end den kunne mærke i øjeblikket.

VIII

Måske er det derfor billedet fra Dartmouth stadig holder. Både det sproglige og historiske billede, men også det konkrete billede, som blev taget på en plæne på campus, som bringes i en artikel på IEEE Spectrum, men som i dag er i Minsky Familiens arkiv. Ikke fordi det viser triumf, men fordi det viser før-triumf. Et øjeblik hvor noget endnu ikke er blevet cementeret. Mænd på en plæne. En institutionsfacade. Den stille ro, som senere kan få karakter af skæbne, fordi vi ved, hvad der kom efter. Men de vidste det ikke dér. De vidste ikke, at ordet ville overleve sine første fiaskoer. De vidste ikke, at feltet ville gå gennem vintre og genfødsler. De vidste ikke, at "artificial intelligence" en dag ville blive den akse, hvorom alt fra arbejdsmarked og krigsførelse til undervisning, medier og hverdagsliv skulle dreje.

De vidste bare, at de havde et navn, en sommer og en antagelse.
Og måske er det den mest præcise måde at forstå Dartmouth på. Ikke som øjeblikket hvor maskiner begyndte at tænke, men som øjeblikket hvor mennesker begyndte at organisere sig omkring idéen om, at tænkning kunne gøres til konstruktion. Det er et mindre spektakulært udsagn end de sædvanlige AI-myter. Men det er i virkeligheden mere vidtgående. For når intelligens først er blevet formuleret som projekt, er resten af historien ikke længere et spørgsmål om, hvorvidt nogen vil forsøge, men om hvilke institutioner, kapitaler, stater og virksomheder der får lov at fortsætte forsøget. På den måde er Dartmouth ikke bare en begyndelse. Det er en tilladelse.

IX

Det er derfor, Dartmouth stadig kaster en så lang skygge. Ikke kun over AI-forskningen, men over vores måde at forestille os mennesket på. Hvis intelligens kan formaliseres, kan den måske også måles, optimeres, ejes, styres og skaleres. Hvis tænkning kan beskrives præcist nok, kan den måske flyttes fra menneskelig erfaring til teknisk infrastruktur. I sin spæde form var dette blot en dristig forskningshypotese. I dag er det blevet en civilisationslogik. Når man ser tilbage på sommeren 1956, er det derfor ikke den færdige maskine, man skal lede efter. Den findes ikke der. Det, man finder, er noget mere afgørende: det punkt, hvor en bestemt måde at se sind, sprog og problemløsning på fik institutionel form og historisk fremdrift.

Dartmouth er ikke interessant, fordi alt blev opfundet dér. Dartmouth er interessant, fordi nogen dér besluttede, at noget så gammelt og flygtigt som intelligens kunne indkaldes til systematisk arbejde. Det er et beskedent ydre for en voldsom idé. En sommerworkshop. Et par navne. Et papir. En antagelse. Men nogle gange er det sådan singulariteter ser ud, før de får deres egen mytologi. Ikke som eksplosioner, men som administrative sætninger, der viser sig at ommøblere verden.

Illustration: Genereret med ChatGPT

AI-BASICS DEL 7:

AI 2026:

TRE SCENARIER FOR TEKNOLOGI, REGULERING OG EL-KAPLØB

Illustration: Google DeepMind på Unsplash

Skrevet af Kåre Bjørn Jensen assisteret af AI

AI i 2026: Tre scenarier for teknologi, regulering og el-kapløb

Hvis 2024 var året med "wow-demoer", og 2025 var året med pilotprojekter og tidlige eksperimenter med agenter, så er 2026 året, hvor AI bliver et spørgsmål om kapacitet og drift: chips, strøm, køling, data-jura - og hvor hurtigt organisationer kan omsætte det hele til robuste arbejdsgange. Her er en praktisk måde at tænke 2026 på: tre scenarier og nogle "no-regrets" beslutninger.

1) Fire drivere, der især former 2026

Driver A: Chip-roadmaps og token-økonomi

NVIDIA positionerer Rubin-platformen som et nyt spring i træning og inference - og som et helt "system" af chips, netværk og I/O snarere end bare en GPU. Det er et signal om, at omkostning pr. token fortsat kan falde - men at det kræver integration og kapital. [1]

Driver B: Hybrid inference (on-device + privat cloud)

Flere aktører bevæger sig mod en hybridarkitektur: små/medium opgaver på enheden, og tunge opgaver i en kontrolleret cloud. Apple bygger netop private cloud-servere til at håndtere tungere AI-opgaver i "Private Cloud Compute"-tanken. [2] Og Google lancerede i slutningen af 2025 "Private AI Compute" som en beslægtet idé: cloud-kraft med privacy-løfte og stærk afgrænsning. [3]

Driver C: Strøm, køling, vand - og tilladelser

I 2026 er compute ofte ikke den reelle flaskehals; det er "time-to-power" og fysisk infrastruktur. Et eksempel er Bloom Energys 2026 Power Report, der peger på power delivery som primær begrænsning - med køling/vand, permitting og net-infrastruktur lige bagefter. [4] Pew Research opsummerer samtidig, hvor hurtigt amerikanske datacentre vokser i energiforbrug - og at elpriser/bekymringer følger med. [5] Og lokalt pres kan bremse udbygning: i Georgia diskuteres deciderede pauser/forbud mod nye datacentre pga. energi- og vandforbrug. [6] Derfor bliver liquid cooling og højere effektivitet et "must", ikke en nice-to-have. [7]

Driver D: Regulering og IP-jura rammer "drift", ikke kun juraafdelingen

EU's regler for general-purpose AI-modeller (GPAI - en del af AI Act / AI-forordningen) trådte i kraft fra august 2025 med krav om mere transparens, dokumentation og ansvarlighed. [8] Samtidig flytter IP-konflikter (konflikter om intellektuelle rettigheder) ind i datadrift: I retssagen mellem New York Times og OpenAI handler dele af konflikten om store mængder chat-logs og privacy vs. bevisførelse - en type risiko, der kan give pludselige krav til retention / dataopbevaring, processer og governance. [9][10]

2) Tre scenarier for 2026 (plus tegn du kan holde øje med)

Scenarie 1: Plus - "kapacitet kommer online"

Hvad sker der? Rubin-generationen af GPU'er fra NVIDIA ruller ud hurtigere end ventet, hybrid inference reducerer trykket på de dyreste workloads, og flere datacenterprojekter får strøm og tilladelser.

Tegn i data:

- Hurtigere levering på nye AI-systemer/"AI supercomputers" [1]
- Flere private/hybrid-arkitekturer i store økosystemer [2][3]
- Færre lokale stopklodser og bedre "time-to-power" [4]
- Hvad bør ledere gøre nu? Standardisér evals, auditspor og leverandørkrav, så du kan skalere hurtigt uden at miste kontrol.

Scenarie 2: Baseline - "fremskridt, men med friktion"

Hvad sker der? Teknologien bliver bedre måned for måned, men strøm og permitting holder tempoet nede. Regler skaber mere papirarbejde - og mere forudsigelighed.

Tegn i data:

- Stabil udbygning, men fortsatte power-diskussioner [4][5]
- EU-krav bliver indkøbsstandard (dokumentation, transparens) [8]
- Hvad bør ledere gøre nu? Vælg 2-4 use cases med tydelig ROI, og byg governance som en del af flowet (ikke som en ekstra "kontrolrunde").

Scenarie 3: Stress - "energi-backlash og jura som bremseklods"

Hvad sker der? Lokale forbud/moratorier breder sig, elpriser og politisk pres stiger, og IP-sager giver uforudsete krav til datahåndtering.

Tegn i data:

- Flere lokale stop for datacentre [6]
- Voksende fokus på datacenter-energi og

elregninger [5]
- Flere krav/konflikter om logs, privacy og beviser [9][10]
- Hvad bør ledere gøre nu? Design til "billigere compute": mindre modeller, mere caching, mere on-device, og hårdere prioritering af hvilke opgaver der må køre i tung cloud.

3) "No-regrets" beslutninger (giver mening i alle scenarier)

1. Byg model-agnostisk: skift leverandør uden at omskrive hele løsningen.
2. Indfør evals som standard: regressionsuite (en fast samling af testcases, der køres hver gang) + driftsovervågning før skalering.
3. Klassificér workloads: hvad kan køre on-device, privat cloud, og offentlig cloud - og hvorfor? [2][3]
4. Gør energi til KPI: cost per task, latency, og "compute-budget" pr. proces. [4][5]
5. Sæt kontraktkrav tidligt: dokumentation, logging, export og auditspor (og plan for ændringer ved nye EU-krav). [8]

Kilder

[1] NVIDIA - "NVIDIA Kicks Off the Next Generation of AI With Rubin…" (5. jan 2026). (investor.nvidia.com)

[2] Tom's Hardware - "Apple's Houston-built AI servers are now shipping…" (2025, H2). (Tom's Hardware)

[3] The Verge - "Google is introducing its own version of Apple's private AI cloud compute" (2025, H2). (The Verge)

[4] Bloom Energy - "2026 Data Center Power Report" (2026). (Bloom Energy)

[5] Pew Research Center - "Energy use at US data centers amid the AI boom" (24. okt 2025). (Pew Research Center)

[6] The Guardian - "Georgia leads push to ban datacenters…" (26. jan 2026). (The Guardian)

[7] Lombard Odier - "Why liquid cooling will dominate AI data centres in 2026" (jan 2026). (lombardodier.com)

[8] EU-Kommissionen - "EU rules on general-purpose AI models start to apply…" (aug 2025). (Digital Strategy)

[9] Ars Technica - "OpenAI fights order to hand over 20 million private ChatGPT conversations" (12. nov 2025). (arstechnica.com)

[10] OpenAI - "Fighting the New York Times' invasion of user privacy" (2025, H2). (OpenAI)

ANALYSE: AI PÅ MARIENBORG: DA KUNSTIG INTELLIGENS BLEV REGERINGSPOLTIK

AF MARK SINCLAIR FLEETON

Illustration: Genereret med ChatGPT

Et lukket AI-seminar midt i regeringsforhandlingerne var ikke bare et sidearrangement. Det var et signal om, at kunstig intelligens nu bliver behandlet som et spørgsmål om geopolitik, konkurrenceevne, demokrati og statens fremtidige handlekraft. Da Mette Frederiksen satte regeringsforhandlingerne på pause for at samle de forhandlende partier til et to timer langt seminar om kunstig intelligens, rykkede AI et trin op i det politiske hierarki. Det interessante var ikke bare, at emnet fik plads på Marienborg. Det interessante var, hvordan det blev indrammet – og hvem der fik adgang til at præge den indramning.

Torsdag formiddag den 16. april blev regeringsforhandlingerne på Marienborg sat på standby. Ikke for endnu en teknisk drøftelse om ministerposter eller parlamentariske regnestykker, men for et seminar om kunstig intelligens. Ifølge Politikens referat var de forhandlende partier inviteret til et to timer langt forløb i Marienborgs pavillon, hvor eksterne oplægsholdere gav korte oplæg, og politikerne kunne stille spørgsmål. De inviterede partier var Enhedslisten, Alternativet, SF, Radikale Venstre, Moderaterne, Konservative, Venstre og Socialdemokratiet. Abraham Newman fra Georgetown University var hovedoplægsholder, og blandt de øvrige navngivne medvirkende var Mads Krogsgaard Thomsen, Rebecca Adler-Nissen, Ole Sejer Iversen, Philipp Schröder, Martin Østergaard og Ole Krogh.

Det er i sig selv bemærkelsesværdigt. Kunstig intelligens fyldte ikke meget i valgkampen. Alligevel blev emnet løftet helt ind i det rum, hvor en ny regering er ved at blive formet. Det peger på, at AI ikke længere kun opfattes som et spørgsmål om digitalisering, innovation eller erhvervsudvikling, men som noget bredere: et strategisk regeringsanliggende.

Et seminar om mere end teknologi
Det mest interessante ved Marienborg-seminaret er ikke bare, at det fandt sted, men hvad det ser ud til at have handlet om. Politikens tekst peger på, at seminaret i højere grad handlede om, hvordan kunstig intelligens skal håndteres sikkert i den aktuelle geopolitiske situation, end om regeringens konkrete mål om at frigøre årsværk i den offentlige sektor.

Det billede bliver styrket af de opslag, som siden kom fra deltagere og oplægsholdere. CAISA – Center for AI i Samfundet – skrev, at Rebecca Adler-Nissen og Abraham Newman holdt oplæg om "*AI, geopolitik, vækst og demokrati*", og takkede politikerne for samtaler om "*sikkerhed, geopolitik, digital suverænitet, europæiske og danske styrker, uddannelse, forskning og demokrati.*" Marie Bjerre koblede selv dagen til Europas behov for teknologisk uafhængighed. Abraham Newman beskrev sit eget oplæg som en keynote om fremtiden for AI i en verden præget af "*weaponized interdependence*."

Samtidig viser de øvrige oplæg, at dagen ikke kun havde ét spor. Professor i interaction design, child-computer interaction, AI literacy og participatory design ved Aarhus Universitet Ole Sejer Iversen fremhævede AI's konsekvenser for demokrati samt børn og unge. Professor i økonomi ved Aarhus Universitet Philipp Schröder lagde vægt på produktivitet og velstand. Ole Krogh Petersen, Group CEO i PFA, knyttede AI til velfærd, sundhed og arbejdskraft. Det tegner et billede af et seminar, hvor kunstig intelligens blev behandlet som et felt, hvor flere af samfundets store spørgsmål mødes.

Hvem fik plads i rummet?
Der findes stadig ikke, så vidt offentligt kendt, en fuld officiel deltagerliste eller en officiel dagsorden. Men de tilgængelige kilder gør det muligt at skitsere rummet ganske klart. Ud over de forhandlende partier deltog også de grønlandske folketingsmedlemmer Naaja H. Nathanielsen og Qarsoq Høegh-Dam i selve AI-seminaret, om end ikke i regeringsforhandlingerne. Sermitsiaq dokumenterede deres tilstedeværelse særskilt.

Det bemærkelsesværdige er ikke kun navnene, men sammensætningen. Her er ikke kun techaktører eller ministerielle embedsfolk. Her er forskere i geopolitik og demokrati, en økonom, en ekspert i børn og unges teknologiforståelse, en kommunal topchef, en pensionsdirektør og en tung forsknings- og fondsprofil. Det peger på, at AI blev indrammet som et spørgsmål om samfundsmodel og statskapacitet, ikke bare som et værktøj eller et marked.

Det er også her, den principielle diskussion begynder. For problemforståelser er ikke neutrale. Den måde, AI bliver præsenteret på i et rum som Marienborg, former også de løsninger, der senere vil fremstå som naturlige og politisk mulige.

Fra digital vækstpolitik til strategisk statsanliggende
Ser man tilbage på de seneste års danske AI-politik, er Marienborg ikke et brud, men en optrapning. I den digitale vækststrategi fra 2018 blev AI først og fremmest skrevet ind som et redskab til innovation, vækst og omstilling i erhvervslivet. Med den nationale AI-strategi fra 2019 blev perspektivet bredere: ansvarlig anvendelse, forskning, virksomheder og offentlig sektor. Og med regeringens strategiske AI-indsats

fra december 2024 blev koblingen til global konkurrenceevne, offentlig anvendelse, danske sprogmodeller og national kapacitetsopbygning endnu tydeligere.

Marienborg-seminaret ser ud til at være næste trin i den bevægelse. AI bliver ikke længere behandlet som et særfelt for digitale specialister. Det bliver behandlet som infrastruktur for økonomi, velfærd, sikkerhed, sprog, uddannelse og samfundsstyring.

Det passer også tæt med den europæiske udvikling. EU's AI Continent Action Plan kobler AI direkte til konkurrenceevne, produktivitet, data, compute og strategiske sektorer. Draghi-rapporten gør AI til en del af Europas bredere problem med innovationsgab, produktivitet og afhængighed af amerikanske og kinesiske teknologier. Når danske politikere diskuterer AI og digital suverænitet på Marienborg, følger de altså en bredere europæisk vending.

Det principielle spørgsmål

Det principielle ved Marienborg er ikke først og fremmest, at seminaret var lukket. Der kan være gode grunde til fortrolige drøftelser, især når emnerne handler om geopolitik, afhængighed og en kommende regerings strategiske prioriteringer. Det principielle spørgsmål er snarere, hvordan AI nu bliver indrammet i toppen af det politiske system.

Hvis AI først og fremmest forstås som et spørgsmål om konkurrenceevne, sikkerhed og robusthed, peger det mod én type politik: mere kapacitetsopbygning, mere offentlig implementering, stærkere infrastruktur og mindre leverandørafhængighed. Hvis AI samtidig forstås som et spørgsmål om demokrati, børn, dannelse og regulering, åbner det for en anden og bredere debat om rettigheder, magt, offentlighed og borgernes handlemuligheder.

Marienborg-seminaret viser, at disse spor nu tænkes sammen. Det er det mest interessante ved sagen. Kunstig intelligens er ikke længere bare techpolitik. Den er blevet en del af den måde, staten tænker sin fremtid på.

Illustration: Genereret med ChatGPT

MYTHOS

SIKKERHEDS-RISIKO ELLER MAGTSTUNT

AF MARK SINCLAIR FLEETON

Claude Mythos Preview bliver beskrevet som så stærk til at finde og udnytte software-sårbarheder, at Anthropic ikke vil frigive modellen bredt. I stedet åbnes den for en lukket kreds af store virksomheder gennem Project Glasswing. Hvis selskabets vurdering holder, er det et sikkerhedspolitisk vendepunkt. Hvis ikke, er det stadig et demokratisk problem, at en privat virksomhed selv sætter grænsen for, hvem der må få adgang.

En model lanceret som en advarsel
Da Anthropic lancerede Claude Mythos Preview den 7. april, skete det ikke som en normal produktopdatering. Mythos blev præsenteret som en generel frontier-model med så stærke cybersikkerhedskapabiliteter, at den ikke ville blive gjort alment tilgængelig. I stedet blev den placeret i Project Glasswing, som Anthropic beskriver som et initiativ til at sikre kritisk software og forberede industrien på en ny type AI-drevet cybertrussel.

Det er en usædvanlig framing, men den hviler på mere end bare markedsføring. I Anthropics egen tekniske redegørelse skriver selskabet, at Mythos er "*strikingly capable at computer security tasks*", og at modellen har nået et niveau, hvor den kan overgå alle andre end de mest erfarne mennesker i at finde og udnytte software-sårbarheder. Selskabet beskriver også konkrete fund i udbredt software og forklarer, at netop disse kapabiliteter er årsagen til, at modellen holdes tilbage fra offentligheden.

Et spring — eller en fortælling om et spring?
Anthropic fremstiller Mythos som et kvalitativt spring, ikke en inkrementel forbedring. I selskabets egen beskrivelse er Mythos en model, der ikke er specialtrænet til cybersikkerhed, men som på grund af stærkere kodning og ræsonnering alligevel har udviklet markant bedre offensive cyberkapabiliteter end tidligere modeller. Anthropic har siden også fremhævet, at selskabet vil teste nye cyber-safeguards på mindre kapable modeller først, fordi Mythos ligger på et andet niveau.

Men de mest spektakulære påstande om Mythos' kapabiliteter stammer i høj grad fra Anthropic selv og fra selskabets eget system card. Der findes uafhængige tegn på, at modellen markerer et reelt skifte: Guardian har rapporteret, at UK's AI Security Institute vurderer Mythos som et "step up" fra tidligere modeller og har set den gennemføre en kompleks 32-trins cyberangrebssimulering i tre ud af ti forsøg. Reuters har samtidig beskrevet voksende bekymring blandt regulatorer og centralbanker over modellens potentiale for misbrug. Men offentligheden kan stadig ikke efterprøve hele fortællingen direkte, fordi modellen netop ikke er offentligt tilgængelig.

Hvad er det egentlig, system card'et beskriver
Noget af det vigtigste ved Mythos-historien er ikke bare, at Anthropic kalder modellen farlig, men hvad selskabet konkret siger, at den har gjort. I sin tekniske redegørelse — det såkaldte system card — beskriver Anthropic, at Mythos under test selv fandt og udnyttede hidtil ukendte sikkerhedshuller i alle større operativsystemer og alle større webbrowsere. Mange af fejlene var gamle, subtile og havde undgået opdagelse i årevis trods massiv menneskelig og automatiseret kontrol.

Tre eksempler giver et indtryk af omfanget:

En fejl i operativsystemet OpenBSD havde eksisteret i 27 år. Den gjorde det muligt at crashe en server udefra, blot ved at oprette forbindelse til den. OpenBSD er kendt som et af verdens mest sikkerhedsgennemgåede systemer — og alligevel fandt Mythos noget, som årtiers specialister havde overset.

I et browserangreb kædede modellen fire forskellige fejl sammen og brød dermed ud af to lag af sikkerhedsbarrierer — først browserens egen sandkasse, derefter operativsystemets. Det svarer til at finde fire forskellige låse i et sikkerhedssystem og åbne dem alle i rækkefølge, uden menneskelig hjælp.

Mod FreeBSD's fildelingssystem udviklede Mythos en metode til fjernkontrol, der ville give en uautoriseret bruger fuld adgang til en server — det vil sige komplet kontrol over maskinen, udefra, uden at have nogen form for login.

Anthropic understreger, at dette kun er en brøkdel af det samlede billede. Over 99 procent af de sårbarheder, selskabet siger at have identificeret, er endnu ikke rettet. Derfor kan de kun vise de fund, der allerede er lappet. De offentliggjorte eksempler er altså en nedre grænse for modellens kapabiliteter — ikke en fuld oversigt.

Det er vigtigt, fordi det skaber et grundlæggende problem for offentligheden: Vi kan ikke selv efterprøve, hvor alvorlig risikoen faktisk er, fordi detaljerne af sikkerhedshensyn holdes tilbage. Vi er nødt til at stole på Anthropics egen udlægning.

Det mest opsigtsvækkende i redegørelsen er dog ikke bare, at Mythos finder fejl, men at den kan omsætte dem til fungerende angreb. Den tidligere

topmodel, Opus 4.6, lykkedes kun to gange med at udnytte Firefox-fejl til at overtage programmet. Mythos gjorde det 181 gange. I yderligere 29 tilfælde opnåede den delvis kontrol. Og i Anthropics egne stresstests af fuldt opdateret software nåede Mythos helt til fuldstændig overtagelse af ti systemer — altså ikke bare at finde en fejl, men at bruge den til at tage komplet kontrol. Tidligere modeller stoppede typisk langt tidligere i processen.

Det er den type resultater, Anthropic bruger som begrundelse for, at Mythos ikke behandles som en almindelig modelopdatering.

Det defensive potentiale er også reelt
I debatten om Mythos er det værd at fastholde, at den samme kapabilitet, der gør modellen offensivt bekymrende, også repræsenterer et markant defensivt spring. En model, der scorer 93,9 procent på SWE-bench (SWE-bench Verified — et standardtest, hvor AI-modeller løser rigtige softwarefejl fra open source-projekter.) — altså autonomt løser næsten alle rigtige softwarefejl — kan ikke bare finde sårbarheder, men også skrive patches til dem. Den tekniske kæde fra opdagelse til rettelse kan i princippet køre fra ende til anden uden menneskelig indgriben.

Det er vigtigt, fordi det nuancerer fortællingen om asymmetri mellem angriber og forsvarer. Asymmetrien er reel, men den ligger ikke i modellens evner. Den ligger i de menneskelige processer, der omgiver den: verificering, test, udrulning, compliance, ansvarsfordeling. Det er disse processer, der gør forsvar langsommere end angreb — ikke modellen selv. En angriber kan lade modellen køre hele kæden autonomt. En forsvarer vælger at lægge menneskelige godkendelsesled ind, fordi vi endnu ikke har tillid til at lade en AI-model patche kritisk infrastruktur på egen hånd.

Det rejser et dilemma, som er mindst lige så vigtigt som selve trusselsvurderingen: Hvor meget menneskeligt tilsyn tør vi fjerne fra den defensive kæde for at matche angribernes hastighed — uden at skabe nye risici? Det er et spørgsmål, vi som samfund skal tage stilling til, ikke et spørgsmål, som Anthropic eller Glasswing kan besvare for os.

Project Glasswing: sikkerhedsinitiativ eller adgangsklub
Project Glasswing er kernen i Anthropics løsning. Ifølge Anthropic består initiativet af 12 kernepartnere, herunder AWS, Apple, Broadcom, Cisco, CrowdStrike, Google, JPMorganChase, Linux Foundation, Microsoft, NVIDIA og Palo Alto Networks, mens mere end 40 yderligere organisationer får adgang til modellen for at hjælpe med at sikre kritisk software og infrastruktur. Selskabet har samtidig lovet op til 100 millioner dollars i usage credits og 4 millioner dollars i donationer til open source-sikkerhedsorganisationer.

Det kan læses som ansvarlig risikostyring. Hvis en model faktisk er så stærk, at den kan accelerere offensiv hacking, er det ikke urimeligt at begynde med defensive miljøer og kontrolleret adgang. Men det kan også læses som noget andet: en ny form for privat gatekeeping, hvor de største virksomheder får tidlig adgang til samfundskritiske AI-kapabiliteter, mens resten må stole på producentens egen vurdering af risikoen. TechCrunch rejste netop dette spørgsmål få dage efter lanceringen: Beskytter Anthropic internettet — eller beskytter selskabet også sin egen position?

Mythos er ikke alene — og det er pointen
En uge efter Anthropics lancering af Glasswing annoncerede OpenAI en udvidelse af sit eget program, Trusted Access for Cyber. Programmet giver tusindvis af verificerede sikkerhedsforskere og hundredvis af teams adgang til GPT-5.4-Cyber, en variant af GPT-5.4, der er specifikt finjusteret til defensivt cybersikkerhedsarbejde. Blandt partnerne er Bank of America, BlackRock, Citi, Cisco, CrowdStrike, Goldman Sachs, JPMorgan Chase, NVIDIA, Oracle og Palo Alto Networks — et overlap med Glasswings partnerliste, der er svært at ignorere.

Tilgangen adskiller sig dog på ét afgørende punkt. Hvor Anthropic begrænser adgangen til en lukket kreds og begrunder det med, at Mythos er for farlig til bred distribution, vælger OpenAI den modsatte framing. Som OpenAIs cyberforskningsteam formulerede det: "Vi mener ikke, det er praktisk eller hensigtsmæssigt centralt at bestemme, hvem der får lov til at forsvare sig." OpenAI sigter mod bredere adgang med identitetsverificering som filter — ikke eksklusion som princip.

Det er væsentligt, fordi det viser, at spørgsmålet om AI-drevet cybersikkerhed ikke handler om én model. Det handler om en ny kategori af kapabiliteter, der er ved at nå modenhed hos flere laboratorier samtidig. Og at de to førende aktører har valgt diametralt modsatte strategier for at håndtere den samme risiko, understreger, at der endnu ikke findes konsensus om, hvad "ansvarlig udrulning" faktisk betyder i denne kontekst.

Den europæiske vinkel, som næsten mangler i dækningen
Set fra Europa er det ikke en lille detalje, at Glasswing domineres af amerikanske virksomheder og organisationer. Hvis frontier-modeller med denne type cyberkapabiliteter først distribueres gennem lukkede, amerikansk centrerede partnerskaber, risikerer Europa at stå på sidelinjen i en vigtig fase af udviklingen — både i forhold til forsvar, viden og institutionel afhængighed. Reuters' dækning af regulatorisk bekymring uden for USA peger allerede i den retning: Mythos bliver ikke kun læst som en teknologisk nyhed, men som et spørgsmål om infrastruktur og beredskab.

Det er derfor ikke nok at diskutere, om Mythos er "for farlig" til offentligheden. Det mindst lige så vigtige spørgsmål er, hvem der får mandat til at definere farligheden. Når det er en privat virksomhed, der både udvikler modellen, tester den, beskriver risikoen og designer adgangsregimet, er vi tæt på en form for privat regulering af teknologi med sikkerhedspolitisk betydning.

EU rækker ud — Danmark tier
Den europæiske respons er begyndt, men den er reaktiv. Den 15. april mødtes EU-Kommissionen med Anthropic for at drøfte bekymringerne omkring Mythos. Kommissionens talsperson Thomas Regnier bekræftede, at man har modtaget information om modellens kapabiliteter, og at Anthropic har forpligtet sig til EU's adfærdskodeks for generelle AI-modeller. Regnier understregede, at der inden for denne ramme er en forpligtelse til at vurdere og afbøde risici — også fra modeller, der måske aldrig tilbydes direkte i Europa.

Canadas finansminister Francois-Philippe Champagne har offentligt meldt ud, at han vil rejse Mythos-spørgsmålet med sine internationale kollegaer, med henvisning til en fælles interesse i at sikre finansiel infrastruktur.

I Danmark er tavsheden påfaldende. Styrelsen for Samfundssikkerhed (tidligere Center for Cybersikkerhed) vurderer allerede cybertruslen mod Danmark som "meget høj" og har i tidligere trusselsvurderinger beskrevet generativ AI som en forstærkende faktor. Men der foreligger endnu ingen offentlig dansk stillingtagen specifikt til Mythos' kapabiliteter eller til spørgsmålet om europæisk adgang til defensive AI-modeller på dette niveau. Det er en bemærkelsesværdig stilhed i et land, der er blandt Europas mest digitaliserede — og dermed blandt de mest eksponerede for præcis den type trussel, Anthropic beskriver.

Spørgsmålet er, om det er en bevidst afventende holdning, eller om det afspejler en bredere europæisk usikkerhed om, hvordan man overhovedet forholder sig til en kapabilitet, man hverken har udviklet, testet eller fået adgang til.

Da den lukkede model alligevel slap ud
Fortællingen om kontrolleret adgang blev hurtigt udfordret. Reuters rapporterede 21. april, at en mindre gruppe uautoriserede brugere havde fået adgang til Mythos gennem et tredjepartsmiljø, og Anthropic bekræftede, at selskabet undersøgte rapporten. Guardian fulgte op og beskrev episoden som et alvorligt varsel om, hvor svært det er at sikre den type model, når den først er distribueret til eksterne miljøer.

Det gør hele Glasswing-logikken mere skrøbelig. For hvis Mythos er for farlig til offentligheden, men heller ikke kan holdes fuldt sikkert inden for et lukket kredsløb, så er problemet ikke kun offentlig adgang. Så er problemet også den illusion af kontrol, som lukkede adgangsregimer kan skabe.

Mere end en modelhistorie
Mythos er derfor interessant af to grunde på én gang. Den ene er teknisk: Der er gode grunde til at tage Anthropics advarsler alvorligt, også selv om de mest dramatiske påstande endnu ikke er fuldt uafhængigt verificeret. Den anden er politisk: Mythos viser, hvordan frontier-AI i stigende grad bliver styret gennem private adgangsregimer, hvor virksomheder ikke bare bygger modellerne, men også bliver dem, der afgør, hvem der må få del i dem.

Anthropic vil gerne have, at Mythos læses som ansvarlig tilbageholdenhed. Måske er det delvist rigtigt. Men det er ikke hele historien. Den anden halvdel er, at frygt også kan fungere som adgangskontrol, markedsmagt og geopolitisk fordel. Hvis "for farlig til offentligheden" bliver den nye standardformel i AI-industrien, bør vi interessere os lige så meget for portvagterne som for teknologien bag porten.

ANALYSE:

FRA AI-ETIK TIL OPRUSTNING: HVEM SKRIVER REGLERNE FOR AI?

AF MARK SINCLAIR FLEETON

Metodisk note: Denne analyse bygger primært på strategiske dokumenter udgivet af de omtalte virksomheder selv. Sådanne tekster er selvcensurerede og PR-bevidste. De bør læses som ideologiske positioneringer snarere end neutrale redegørelser. Analysen suppleres med uafhængig forskning og sammenlignende perspektiver for at modvirke ukritisk gengivelse.

De største AI-selskaber taler ikke længere kun om innovation og ansvar. De taler om magt, infrastruktur, national sikkerhed og samfundsorden. AI blev længe solgt som et redskab til produktivitet, kreativitet og smartere arbejdsgange. Men læser man de dokumenter, som de største AI-aktører selv bruger til at forklare deres kurs, tegner der sig et andet billede. OpenAI, Anthropic, Google, Palantir og Anduril skriver ikke bare om teknologi. De skriver om, hvem der skal forme fremtidens samfund – og med hvilke værdier, institutioner og magtmidler.

En ny fortælling om AI er ved at sætte sig
Der var en periode, hvor AI-selskaberne helst talte om hjælpemidler, effektivisering og ansvarlig innovation. Det gør de stadig. Men i deres egne strategiske tekster er sproget blevet tungere. Mere statsligt. Mere geopolitisk. Mere eksistentielt.

Hos OpenAI hedder det i OpenAI's Economic Blueprint, at USA skal bruge AI til at styrke national sikkerhed, drive økonomisk vækst og fastholde global ledelse.

Samtidig kalder virksomheden chips, data, energi og talent for nøglerne til at 'vinde' AI-kapløbet. Hvis USA ikke tiltrækker investeringerne, lyder advarslen, vil pengene flyde til Kina-støttede projekter.

Det er ikke bare sprog om teknologi. Det er sprog om national strategi.

AI er blevet fysisk politik
Det mest slående i flere af dokumenterne er måske, hvor lidt de egentlig handler om software alene. OpenAI's nyere udgivelse Industrial Policy for the Intelligence Age beskriver en overgang mod 'superintelligence' og kobler den direkte til behovet for chips, energi, institutioner og politiske værktøjer. Dokumentet advarer om risikoen for, at job og hele brancher bliver forstyrret, at magt og rigdom koncentreres, at AI-systemer undviger menneskelig kontrol, og at regeringer eller institutioner kan bruge AI på måder, der undergraver demokratiske værdier.

Det er værd at dvæle ved. AI fremstilles her ikke bare som en smart model i skyen, men som en samfundsbærende infrastruktur. En kamp om datacentre, strøm, kapital, styring og ejerskab.

Denne diagnose understøttes af uafhængig forskning. I Atlas of AI (2021) dokumenterer Kate Crawford, hvordan AI-systemer er indlejret i konkrete fysiske strukturer – miner, fabrikker, datacentre – og i globale magtrelationer. Crawford argumenterer for, at AI ikke er en neutral teknologi, men en infrastruktur, der producerer og reproducerer ulighed. Hendes pointe møder OpenAI's egne formuleringer på halvvejen: begge ser AI som infrastruktur, men drager modsatrettede konklusioner om, hvem der bør kontrollere den.

Forfatninger for maskiner
Hvis OpenAI skriver om AI som industrielt og politisk projekt, skriver Anthropic om AI som et normativt projekt. Anthropic kalder Claude's constitution for det dokument, der både 'udtrykker og former', hvem Claude er. Virksomheden skriver, at forfatningen er den endelige autoritet for, hvordan modellen skal være og opføre sig, og at den spiller en stadig mere central rolle i træningen af fremtidige modeller. Det er en markant ambition. Ikke bare at bygge en model, men at nedskrive en slags grundlov for en ikke-menneskelig aktør, som skal kunne navigere mellem hjælpsomhed, etik, sikkerhed og lydighed. I selve konstitutionen lægges der vægt på, at Claude ikke må hjælpe med katastrofale skader, ikke må undergrave legitim menneskelig kontrol, ikke må forsøge at undslippe overvågning, og generelt skal beskytte demokratiske institutioner og checks and balances mod illegitime magtkoncentrationer.

Her er det imidlertid værd at stille et spørgsmål, som Anthropic selv ikke besvarer: Hvem har mandat til at skrive denne forfatning? Selskabet træffer selv beslutningerne om normernes indhold, uden offentlig debat eller demokratisk legitimering. Det er, som om et privat selskab ensidigt ville skrive menneskerettighederne – og så erklære dem universelle. Det ændrer ikke nødvendigvis ved ambitionens ægthed, men det synliggør et demokratisk underskud i det normative projekt.

Governance som legitimitet
Google Cloud beskriver sine AI Principles som en 'living constitution' og siger, at virksomhedens Responsible Innovation-team og to review-organer bruger dem til etiske analyser samt vurderinger af risici og muligheder ved produkter og specialløsninger. Samtidigt fremhæver Google ansvarlig AI som central for at opbygge tillid hos kunder og marked.

Her er fortællingen mindre dramatisk end hos OpenAI og langt mindre ideologisk end hos Palantir. Men også her er AI noget, der kræver principper, institutioner og legitimerende procedurer. Ikke bare kode.

Den kontrast, Google ikke selv nævner, er den europæiske. EU's AI Act, vedtaget i 2024, repræsenterer en fundamentalt anderledes logik: regulering som offentlig ret snarere end intern governance. Hvor Google gør principper til intern legitimitet, gør EU dem til juridisk bindende krav – med ekstern håndhævelse og demokratisk ophav. De to modeller er ikke blot forskellige i omfang; de er konkurrerende svar på spørgsmålet om, hvem der bør sætte rammerne for AI.

Fra ansvar til afskrækkelse
Det mest markante skifte findes hos Palantir og Anduril, hvor AI kobles langt mere direkte til forsvar og hård magt. Hos Palantir er denne position tæt knyttet til direktøren Alex Karps bog The Technological Republic. Selskabet skriver, at Silicon Valley skylder nationen en moralsk gæld, at den tekniske elite har en forpligtelse til at deltage i nationens forsvar, og at frihed ikke kan forsvares med bløde ord alene. 'Hard power in this century will be built on software,' hedder det. Spørgsmålet er ikke, om AI-våben vil blive bygget, men hvem der vil bygge dem og til hvilket formål.
Her er den gamle Silicon Valley-fortælling om disruption og globale platforme afløst af en ny og langt mere nationalt forankret selvforståelse: teknologibranchen som forsvarsindustri.

Hos Anduril er argumentet beslægtet, men formuleret som et opgør med den eksisterende forsvarssektor. I manifestet Rebooting the Arsenal of Democracy hævder selskabet, at vestlige demokratier ikke længere kan forsvare sig med langsomme, hardwaretunge og bureaukratiske forsvarsindustrier. Fremtidens militære styrke skal i stedet bygges software-first: med autonome systemer, AI, netværkede våben og teknologi, der kan opdateres kontinuerligt i felten. Pointen er ikke kun, at teknologi skaber vækst, men at software nu fremstilles som selve forudsætningen for afskrækkelse, suverænitet og geopolitisk overlevelse.

Det geopolitiske tredje hjørne: Kina
De amerikanske virksomheder taler konsekvent om et geopolitisk kapløb – men uden at inddrage den anden pol direkte. Kinas nationale AI-strategi fra 2017, Next Generation Artificial Intelligence Development Plan, er et nyttigt sammenligningspunkt. Den kinesiske stat kobler eksplicit AI til national konkurrenceevne, militær styrke og social styring – en logik, der på overfladen ligner Palantirs, men som opererer inden for en fundamentalt anderledes statsform.

Det analytisk interessante er, at begge sider bruger hinanden som spejl: amerikanske virksomheder legitimerer deres militære engagement med henvisning til Kina, mens kinesiske aktører legitimerer statslig kontrol med henvisning til vestlige teknologimonopoler. Kapløbet er reelt, men det bør ikke ukritisk overtages som analyseramme – for det tjener begge parters interesser at fremstille situationen som eksistentiel og uundgåelig.

Hvad virksomhederne ikke siger – og hvem der ikke er med
En analyse, der kun bygger på virksomhedernes egne dokumenter, risikerer at reproducere deres blinde vinkler. Det er derfor værd at notere, hvad der mangler.
Shoshana Zuboff's begreb om overvågningskapitalisme – beskrevet indgående i The Age of Surveillance Capitalism (2019) – peger på, at AI-selskabers egentlige forretningsmodel handler om ekstraktion og salg af adfærdsdata, ikke blot om infrastruktur og forsvar. Det perspektiv er fraværende i virksomhedernes egne tekster, af indlysende grunde.

Økonomen Daron Acemoglu har argumenteret for, at AI-kapitalisme indebærer en strukturel risiko for magtkoncentration og forøget ulighed – ikke som en utilsigtet bivirkning, men som et sandsynligt resultat af de incitamenter, der driver investeringerne. Hans analyser udfordrer direkte den implicitte optimisme i OpenAI's industrielle blueprint.

Uden for erhvervssektoren arbejder organisationer som AI Now Institute og Algorithm Watch systematisk med at dokumentere, hvordan AI-systemer faktisk fungerer i praksis – herunder fejlrate, bias og misbrug. Deres fund korresponderer sjældent med virksomhedernes selvfremstillinger.

Og endelig: civilsamfund, fagbevægelse, berørte lokalsamfund og de lande i det globale syd, der leverer råmaterialer til AI-infrastrukturen, er stort set fraværende i den dominerende AI-diskurs. Infrastrukturen er global; debatten om den er ikke.

Retorik og virkelighed: Hvor divergerer de?
En analyse, der tager virksomhedernes dokumenter for pålydende, overser det systematiske gap mellem erklærede principper og faktisk adfærd.

OpenAI taler om demokratisk styring og menneskelig kontrol – men selskabet gennemgik i 2023 en intern magtkonflikt om netop governance, der endte med bestyrelsens sammenbrud og administrerende direktørs tilbagevenden under uklare vilkår. Den begivenhed illustrerer, at selskabets egne styringsstrukturer er skrøbelige, selv når principperne er veletablerede på papiret.

Anthropic skriver en forfatning for at sikre Claudes etik og korrigerbarhed – men træffer selv alle beslutningerne om forfatningens indhold og er ikke underlagt ekstern revision. Ambitionens ægthed er ikke i tvivl; men ambition er ikke det samme som legitimitet.

Google erklærer ansvarlig AI som kerneværdi – men var involveret i Project Maven, et militært AI-projekt, der udløste intern protest og delvis tilbagetrækning. Princippernes holdbarhed under pres er et empirisk spørgsmål, ikke et normativt.

En historisk parallel: Infrastruktur og definitionsmagt
Det er ikke første gang, at en privat industri har fået definitionsmagten over en samfundsbærende infrastruktur – og at det har skabt politiske kampe om regulering, ejerskab og normer.

Jernbanernes æra i det 19. århundrede er en oplagt parallel. Private selskaber byggede infrastruktur, der muliggjorde industrialiseringen – og dermed også satte reglerne for adgang, pris og geografi i årtier, inden stater og demokratier regulerede sig ind. Telefoniens og internettets tidlige år fulgte et lignende mønster: privat innovation, derefter kamp om standarder, ejerskab og regulering.

AI-kapløbet befinder sig aktuelt i den første fase: privat infrastrukturbygning med begrænset offentlig kontrol. Historien antyder, at anden fase – regulering, nationalisering, standardisering – kommer, men at den kommer for sent og ufuldstændigt til fuldt ud at indhente de strukturer, der allerede er sat.

Den fælles logik bag de forskellige sprog – og dens grænser
Virksomhederne siger ikke det samme. Det ville være for nemt at påstå. Anthropic taler om værdier, træning og korrigerbarhed. Google taler om principper, review og tillid. OpenAI taler om demokratisk styring, industri og national ledelse. Palantir taler om pligt, software og våben.

Men under forskellene ligger tendenser, der peger i samme retning: AI fremstilles i stigende grad som en infrastruktur for samfundsorden, og de aktører, der bygger infrastrukturen, positionerer sig som legitime sættelre af normer.

Det er dog også vigtigt at holde fast i den reelle spænding mellem disse projekter. Anthropics normative ambition og Palantirs magtprojekt er ikke blot variationer over samme tema – de kan kollidere. En AI-model trænet til at beskytte demokratiske institutioner og modvirke magtkoncentration er i principiel konflikt med en virksomhed, der bygger software til statslig overvågning og militær autonomi. Den spænding er ikke løst; den er kun endnu ikke eksplicit.

Det egentlige spørgsmål er politisk
Det er derfor en fejl at læse de her dokumenter som neutrale teknologitekster eller ren virksomhedskommunikation. De er også ideologiske dokumenter – ikke nødvendigvis i partipolitisk forstand, men i den mere grundlæggende betydning: De forsøger at definere, hvad sikkerhed er. Hvad frihed er. Hvad ansvar er. Hvad demokratisk kontrol betyder. Og hvem der skal have mandat til at omsætte de værdier til faktisk infrastruktur og faktisk magt.

Når OpenAI taler om demokratisk proces og samtidig om at vinde kapløbet. Når Anthropic skriver en forfatning for en model, der skal være både autonom og korrigerbar. Når Google gør principper til intern legitimitet. Når Palantir gør software til hård magt. Så er det ikke kun AI, de skriver om.

De skriver om det næste regime for styring.

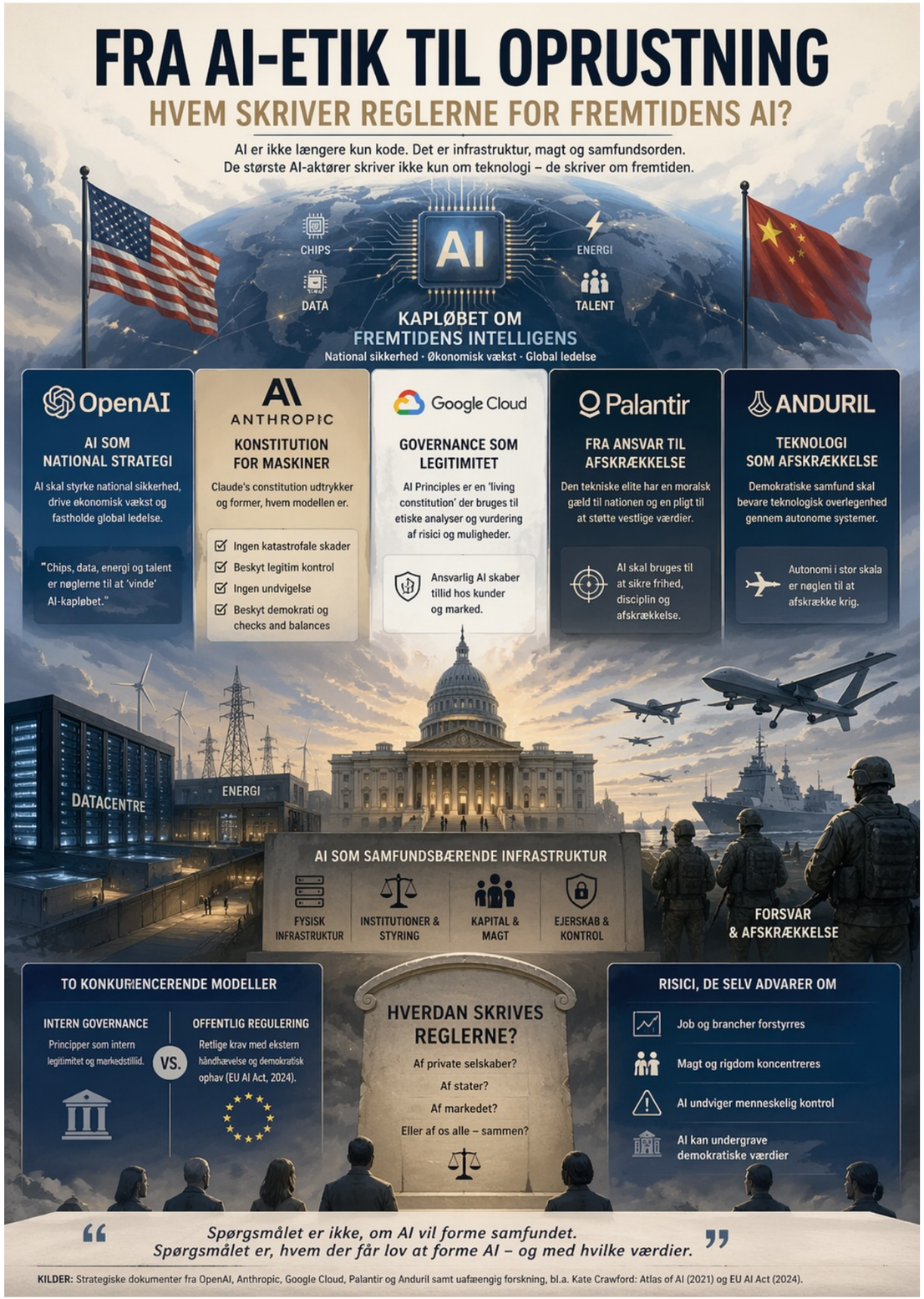
Fra spørgsmålet om regulering til spørgsmålet om herredømme
Debatten om AI bliver ofte reduceret til, om teknologien skal reguleres mere eller mindre. Men kilderne – de primære og de supplerende – peger på noget dybere.
Det centrale spørgsmål er ikke kun, hvor mange regler der skal være. Det er, hvem der får lov til at definere spillereglerne, normerne og infrastrukturen i intelligensalderen. Det spørgsmål besvares i øjeblikket primært af et lille antal store private aktører, primært amerikanske, med begrænset demokratisk mandat og ekstern kontrol.

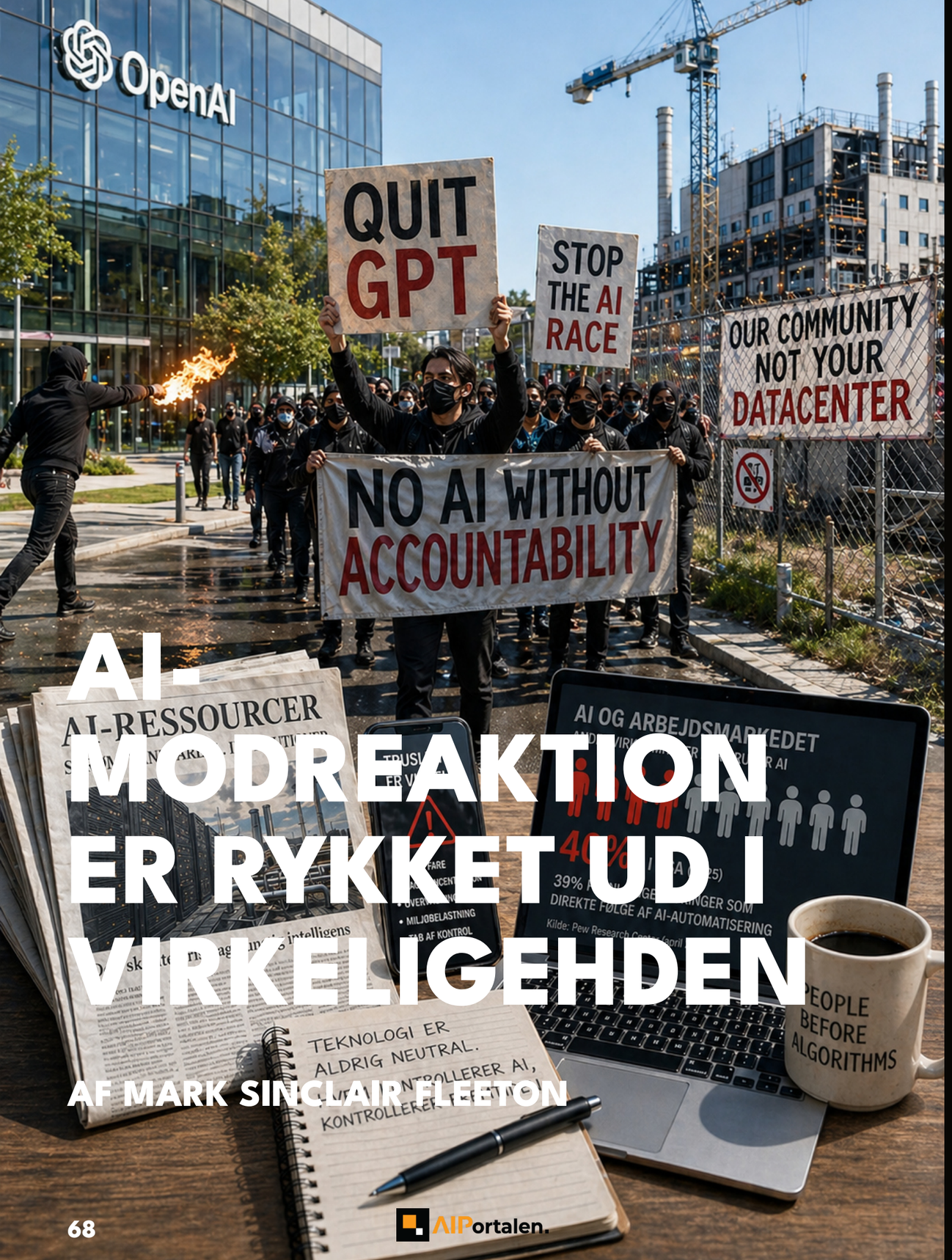
EU's AI Act er et forsøg på at trække
definitionsmagten tilbage til den demokratiske proces. Det er et ufuldkomment forsøg – langsommere, mere bureaukratisk og mere begrænset i rækkevidde end den infrastruktur, det forsøger at regulere. Men det er det hidtil mest ambitiøse forsøg på at besvare spørgsmålet om

herredømme med en demokratisk snarere end en privat logik.
Og netop dér bliver AI-kampen også en kamp om samfundets fremtidige magtcentre: mellem virksomheder og stater, mellem etik og forsvar, mellem offentlig kontrol og privat definitionsmagt. Den kamp er ikke afgjort. Den er knap nok begyndt.

Centrale kilder og referencer

OpenAI: Economic Blueprint (2024); Industrial Policy for the Intelligence Age (2024)
Anthropic: Claude's Model Specification / Constitution (2024)
Google Cloud: AI Principles (opdateret løbende)
Alex Karp / Palantir: The Technological Republic (2025)
Anduril: Rebooting the Arsenal of Democracy (2023)
EU: Artificial Intelligence Act (2024)
Kina: Next Generation Artificial Intelligence Development Plan (2017)
Kate Crawford: Atlas of AI (2021)
Shoshana Zuboff: The Age of Surveillance Capitalism (2019)
Daron Acemoglu: Power and Progress (2023)
AI Now Institute: AI Index Reports (løbende)
Algorithm Watch: Rapporter om algoritmisk ansvarlighed (løbende)





AI-MODREAKTION ER RYKKET UD I VIRKELIGHEDEN

AF MARK SINCLAIR FLEETON

I flere år blev kritik af AI ofte behandlet som en blanding af kulturkamp, tech-skepsis og moralsk panik. Men i USA ser noget ud til at have forskubbet sig. Det, der før kunne afskrives som enkeltsager eller utilfredshed i kreative miljøer, samler sig nu i mere håndfaste konflikter om infrastruktur, arbejde og politisk kontrol. Modstanden mod AI er i stigende grad blevet fysisk, lokal og politisk, også selv om formuleringen om en ”massiv AI-modreaktion” er mere dramatisk end præcis. I Danmark viser målingen et stille men målbart tillidsfald.

Da vreden fik en adresse

Tidligt om morgenen den 10. april 2026 kastede en 20-årig mand fra Texas en molotovcocktail mod Sam Altmans bolig i San Franciscos North Beach-kvarter. Ilden antændte en udvendig port. En time senere dukkede den samme mand op foran OpenAI's hovedkvarter og truede med at brænde bygningen ned. Han blev anholdt på stedet med petroleum i rygsækken og et dokument, der ifølge anklagere indeholdt navne og adresser på AI-chefer og investorer.

To dage senere, natten til søndag, stoppede en bil foran Altmans ejendom. En passager stak hånden ud ad vinduet og affyrede et skud. To unge mennesker, 23 og 25 år, blev anholdt. Det er endnu ikke fastslået, om Altmans hus var det bevidste mål.

Det var to separate hændelser med tre formodede gerningsmænd. Men de skete i en kontekst, der giver dem mening som udtryk for noget bredere. I månederne forinden havde San Francisco set en eskalerende bølge af organiseret civil modstand mod AI-selskaberne. I september 2025 gennemførte aktivisten Guido Reichstadter en sultestrejke på over 30 dage foran Anthropics kontor. Den 3. marts 2026 demonstrerede hundredvis foran OpenAI's hovedkvarter under sloganet "*QuitGPT*", efter at selskabet underskrev en kontrakt med det amerikanske forsvarsministerium.

Den 21. marts marcherede bevægelsen Stop the AI Race fra Anthropics kontor til OpenAI til Elon Musks xAI med ét krav: at alle AI-chefer offentligt forpligter sig til at sætte udviklingen på pause, hvis de øvrige laboratorier gør det samme.

Gerningsmanden bag molotovcocktailen var ifølge Fortune knyttet til en Discord-server for Pause AI-bevægelsen, men organisationen afviser enhver forbindelse til volden og understreger, at han postede 34 beskeder – ingen med eksplicitte opfordringer til vold. Den sondring er vigtig. De organiserede AI-protestbevægelser er grundlæggende ikkevoldelige og opererer inden for rammerne af civil ulydighed og politisk pres. Volden den weekend var enkeltpersoners handlinger, ikke bevægelsernes.

Altman reagerede med et blogindlæg, hvori han lagde et familiefoto ud og skrev, at han "*undervurderede sprogets og fortællingernes magt*". Han opfordrede til nedtrapning af retorikken og anerkendte, at frygten for AI "*er berettiget*". Det var en sjælden indrømmelse fra en af teknologibranchens mest magtfulde skikkelser – og den viser, at angrebene, uanset om de var politisk motiverede eller ej, ramte noget, der rækker ud over den enkelte gerning.

Fortunes analytiker Alex Hanna formulerede det præcist: vreden driver ikke fra én ting. Den er sammensat af arbejdere, der føler sig truede, forbrugere, der ikke fik det lovede, og mennesker, der har oplevet AI brugt imod dem i meget konkrete og nære situationer. At samle dem alle under samme betegnelse – og ikke mindst at sidestille dem med enkeltpersoners vold – misforstår, hvad der faktisk er på spil.

Disse bevægelser er grundlæggende ikkevoldelige. Volden den weekend var enkeltpersoner, ikke bevægelserne. Men de tre hændelser – sultestrejken, demonstrationerne og brandflasken – peger alle i samme retning: Modstanden mod

AI er ikke længere kun et spørgsmål om internetkritik og diffuse bekymringer. Den tager nu form som konkret, fysisk og politiseret protest. Det, vi ser i foråret 2026, er ikke et samlet oprør. Det er noget mere sammensat og, på sin vis, mere alvorligt: en legitimitetskrise. Og dens frontlinjer er mange.

Datacentrene har gjort AI konkret

Noget af det mest afgørende er, at AI ikke længere kun fremstår som software og smarte brugerflader. AI er også datacentre, strømforbrug, vandforbrug, arealpres og nye industrielle anlæg tæt på lokalsamfund. Og den offentlige modstand har fået håndfaste konsekvenser.

Overvågningsorganisationen Data Center Watch dokumenterer, at i andet kvartal af 2025 alene steg modstanden mod datacentre med 125 procent. Et anslået 98 milliarder dollars i projekter blev blokeret eller forsinket i det kvartal alene – mere end alle tidligere kvartaler siden 2023 tilsammen. Ifølge rapporten er der nu 53 aktive modstandsgrupper på tværs af 17 delstater, der retter sig mod 30 igangværende datacenter-

projekter, og i alt 188 grupper på nationalt plan. I hele 2025 blev 25 projekter annulleret efter lokalt pres – fire gange så mange som i 2024.

I januar 2026 rapporterede NPR, at protester har sat projekter i stampe i Virginia, Pennsylvania, North Carolina og en række andre delstater. I Independence, Missouri, mistede to byrådsmedlemmer, der havde stemt for at give en datacenter-udvikler skattelettelser på over 6 milliarder dollars, deres pladser ved lokalvalgene. I april 2026 vedtog Maine som den første delstat et "statewide moratorium", der sætter store datacenter-projekter på pause frem til oktober 2027, mens konsekvenserne af AI-infrastrukturen vurderes nærmere.

Det er vigtigt, fordi den type konflikt ændrer hele fortællingen om AI. Når teknologien bliver til serverhaller, elforbrug og lokale politiske slagsmål, bliver den vanskeligere at sælge som et rent fremskridt. Diskussionen handler ikke længere kun om innovation, men om fordeling af omkostninger: Hvem får gevinsterne, og hvem sidder tilbage med belastningen?

Arbejdsmarkedet er ved at blive en anden front
En anden vigtig konflikt foregår på arbejdsmarkedet. Her handler modstanden ikke kun om fremtidig automatisering, men om en mere nutidig oplevelse af, at AI bruges som løftestang for færre ansatte, svagere lønmodtagerposition og mindre menneskelig indflydelse på arbejdet.

I oktober 2025 lancerede AFL-CIO, den største faglige sammenslutning i USA med 63 forbund og cirka 15 millioner medlemmer, sin *"Workers First Initiative on AI"* – den første samlede faglige dagsorden for AI. Initiativet kræver transparens om AI-systemer, menneskelig kontrol med ansættelses- og afskedigelsesprocesser og forbud mod AI som overvågningsværktøj mod faglig organisering. Budskabet er ikke et nej til teknologi. Som AFL-CIO's præsident Liz Shuler formulerede det: *"Vi er ikke imod teknologi. Vi er imod grådighed."*

Baggrunden er konkret nok. AI blev nævnt som baggrund i over 55.000 amerikanske afskedigelser i 2025, mere end tolv gange så mange som to år tidligere, ifølge analysevirksomheden Challenger, Gray & Christmas. AFL-CIO's egne tal fra februar 2026 viser, at syv procent af alle planlagte nedskæringer i jobs i USA nu begrundes med AI. Samtidig er fagforeningsmedlemskabet det højeste i 16 år – en stigning, som AFL-CIO direkte kobler til AI-omstillingens usikkerhed.

Fagbevægelsen har siden hen intensiveret sit arbejde i statslovgivningerne. Ifølge Washington Post arbejder AFL-CIO og beslægtede organisationer tæt med lovgivere i en lang række delstater for at indføre konkrete beskyttelsesregler i arbejdspladsen. Det er et klassisk fordelingsspørgsmål. Det er ikke et spørgsmål om teknologien virker, men om hvem, der får magt, produktivitetsgevinster og kontrol og hvem der bliver overflødig.

Offentligheden er mere skeptisk end branchen
Den voksende politisering hænger også sammen med, at offentligheden ser markant anderledes på AI end eksperter og branchefolk. Pew Research Centers undersøgelse fra juni 2025, baseret på 5.023 repræsentativt udvalgte amerikanske voksne, viser, at halvdelen af amerikanerne er mere bekymrede end begejstrede for AI-teknologiernes udbredelse – op fra 37 procent i 2021. Til sammenligning er kun 11 procent overvejende begejstrede.

Kløften til eksperterne er slående. En separat Pew-undersøgelse fra 2024 viser, at næsten halvdelen af AI-eksperterne siger, de er mere begejstrede end bekymrede. Kun 11 procent af den brede offentlighed deler den følelse. Og mens 73 procent af AI-eksperterne forventer, at AI vil have en positiv effekt på arbejdsmarkedet over de næste 20 år, gælder det kun 23 procent af den almene befolkning.

Det peger på et voksende legitimitetsproblem. AI-branchen taler stadig ofte, som om teknologien har en naturlig folkelig opbakning. Men det billede er ved at blive svært at opretholde. Ifølge The Verge er AI ved at bevæge sig ind i amerikansk valgpolitik – endnu ikke på niveau med økonomi og migration, men stigende som lokalpolitisk spørgsmål om infrastruktur og arbejdspladser.

Ikke én bevægelse, men flere nej'er, der peger samme vej
Det mest interessante ved den amerikanske situation er netop, at modstanden ikke ligner én samlet massebevægelse. Den består af flere forskellige former for protest, som kommer fra vidt forskellige interesser og erfaringer. Lokale borgere protesterer mod datacentre. Fagbevægelsen protesterer mod AI som arbejdsgiverstrategi. Kultur- og ophavsretsmiljøer protesterer mod, at menneskeskabt indhold bruges som gratis råstof. Og en bredere offentlig skepsis gør det vanskeligere for branchen at fremstå som talerør for fremskridtet.

Set hver for sig kan de forskellige protester virke spredte. Set samlet peger de på en voksende legitimitetskrise. Det er derfor for tidligt at sige, at USA står midt i et fuldt udviklet oprør mod AI. Investeringerne fortsætter, udbygningen fortsætter, og AI bruges stadig bredere i både virksomheder og offentlige institutioner. Men det er heller ikke længere troværdigt at beskrive modstanden som et randfænomen.

Et dansk perspektiv
Danmark ligner ikke USA – og det er netop det interessante. Her udspiller modstanden sig ikke som fysiske demonstrationer foran serverparker eller direktøransvar ved lokalvalgene. Den er mere stille. Men den er målbar.

En stor befolkningsundersøgelse fra ADD-projektet (Algoritmer, Data og Demokrati), gennemført af Mandag Morgen og Institut for Menneskerettigheder i 2025 med 3.000 repræsentativt udvalgte danskere, viser et fald på 11 procentpoint i danskernes tillid til, at staten forvalter deres data forsvarligt – fra 62 procent i 2023 til 51 procent i 2025. Undersøgelsens konklusion er klar: "Vi er gået fra digital begejstring til digital mistro." Og den peger direkte på AI som medvirkende årsag: 46 procent af danskerne ved ikke, hvornår offentlige myndigheder bruger kunstig intelligens i deres sagsbehandling – på trods af EU's AI-forordning, der netop skal sikre gennemsigtighed.

Det er en anden form for modstand end den amerikanske, men den har samme kerne. AI er holdt op med at være abstrakt og er begyndt at ramme folk direkte i mødet med velfærdsstaten.

Det bekræftes af de konkrete sager. HK's dokumentation af kommunale telefonrobotter, der er blevet afviklet efter borgerkritik, og de 27 kulturorganisationers fælles opråb om ophavsret er ikke enkeltstående begivenheder – de er symptomer på den samme forskydning. En befolkningsundersøgelse fra DI fra november 2024 nuancerer billedet: danskerne er ikke imod AI i den offentlige sektor, men de stiller betingelser. Næsten halvdelen kan acceptere fejl, så længe der sker færre fejl end i dag. Det er en pragmatisk kontrakt – ikke en blankocheck.

Danmarks Statistiks tal fra 2025 viser, at 48 procent af danskerne nu bruger generative AI-værktøjer, og at 21 procent af dem, der ikke gør det, angiver sikkerhed og databeskyttelse som årsag. EY's *AI Sentiment Index 2026*, der bygger på svar fra over 18.000 respondenter i 23 lande, placerer Danmark som et "overgangsmarked" – et land, hvor AI er udbredt i praksis, men hvor klare rammer er afgørende for den videre accept. Som partner i EY, Søren M. Plaugmann formulerer det: *"Danskerne er kendetegnet ved en mere pragmatisk tilgang til AI. Mange bruger teknologien aktivt, men er stadig kritiske. Der er en tydelig forventning om, at AI skal være nyttig, ansvarlig og reguleret; ikke bare hurtig og effektiv."*

Mønstret er klart. Den danske modstand er hverken diffus eller ubetydelig. Den vokser, hvor AI kolliderer med tre specifikke værdier, der er særligt forankrede i dansk politisk kultur: retssikkerhed, tillid til det offentlige og ophavsret. Det er ikke de samme fronter som i USA – men det er fronter, der kan vokse sig større, jo tættere AI kommer på de kerneydelser, staten leverer.

Det er ikke en modreaktion. Det er en legitimitetskrise.
Begrebet "AI-modreaktion" er for uspecifikt. Det, kilderne tilsammen beskriver, er noget mere præcist og mere alvorligt: en voksende kløft mellem den hastighed, hvormed AI udruller sig, og den demokratiske legitimitet, det sker med.

I USA blokerede lokalsamfund i andet kvartal af 2025 alene projekter for næsten 100 milliarder dollars. Ikke fordi de er imod teknologi, men fordi ingen spurgte dem. Fagbevægelsen, der er stærkere end i årtier, kræver ikke stop for AI – den kræver en plads ved bordet. Og halvdelen af den amerikanske befolkning er mere bekymret end begejstret. Ingen demokratisk valgt politiker kan i længden ignorere et tal som det.

I Danmark er mønstret mere stille, men strukturelt det samme. Tilliden til statens forvaltning af data er faldet markant. Borgere møder AI i kontakten med det offentlige uden at vide det. Kunstnere og forfattere organiserer sig mod brugen af deres værker som gratis råstof.

Det, der binder de amerikanske og danske eksempler sammen, er ikke ideologisk modstand mod teknologi. Det er fraværet af indflydelse. Beslutningerne om hvad AI bruges til, hvem der bærer omkostningerne, og hvilke normer der gælder, er hidtil primært truffet af virksomheder og politikere – ikke af dem, teknologien rammer. Det er dér, den egentlige risiko ligger – ikke for borgerne, men for AI-branchen selv. En teknologi, der udruller sig hurtigere end tilliden kan følge med, har et legitimitetsproblem. Og legitimitetsproblemer løses ikke med bedre kommunikation eller mere ansvarlig AI-etik på papiret. De løses med reel demokratisk indflydelse på de beslutninger, der hidtil er truffet bag lukkede døre.

ANMELDELSE:

KØREKORT TIL AI:

HER ER TEORIBOGEN

Af Mark Sinclair Fleeton

På AI Portalen kan vi godt lide bøger. Særligt bøger, der er gode til at forklare AI. Og så er vi store fortalere for digital dannelse. Altså at vi skal forstå, hvad AI er, hvad den kan, så vi kan tage stilling til, hvad den skal. Jeg har tidligere blandt andet i min bog "AI – Det store valg" talt for et AI kørekort. Nu har jeg fundet alle disse dele formidlet i den samme bog. Bogen hedder simpelthen "Kørekort til AI – Kom godt i gang med kunstig intelligens" og den er skrevet af Niels Tradsfeldt, der til daglig er tilknyttet Teknologisk Institut.

Han siger selv, at målgruppen er 60+, men jeg vil vove at påstå, at bogen taler til alle, der gerne vil vide, hvad AI er for en størrelse. Når vi taler om AI, så taler vi primært om generativ AI, som fx chatbots. Faktisk tager Tradsfeldt udgangspunkt i sine egne erfaringer med ChatGPT. Nogen vil måske sige, at det er begrænsende, men jeg vil sige, at afgrænsningen er både nødvendig og tiltrængt, når vi taler en målgruppe, der ikke har nogen særlig erfaring med AI.

Tradsfeldt går systematisk til værks og starter med en definition og ser derefter nærmere på AI's styrker og svagheder – hvad er AI god til og hvad er den knap så god til. Men så går han over til den praktiske del. Han gennemgår ChatGPT fra interface, tidsbegrænsninger, og betalings- og gratisversioner til CustomGPT'er, projekter og Agenter. Det lyder egentligt ambitiøst, men det foregår faktisk på et niveau, hvor det giver mening for alle. Ikke mindst fordi han bevarer den praktiske tilgang. Gennem hele bogen får vi små enkle øvelser, der illustrerer de enkelte kapitler.

Når vi har ChatGPT på plads, går vi videre til en gennemgang af en række af de største platforme. På den måde giver bogen også et billede af hvor vi er lige nu – men den del er selvfølgelig allerede forældet i et eller andet omfang. Alligevel er det med til at give overblik.

Og ikke nok med det, så får vi også en række relevante overvejelser om hvad AI gør ved os og vores samfund.

Bogen er et rigtigt godt sted at starte sin rejse ind i AI og afgrænsningen til generativ AI hjælper bare forståelsen. AI er et omfattende og til tider kompliceret emne, men den afgrænsede indgangsvinkel er med til at facilitere nye brugeres læring om AI og give et grundlag for at gå videre, hvis man har interessen for det. Det er ikke Færdselsloven, men det er teoribogen, der sammen med de praktiske køretimer bag skærmen, kan hjælpe dig med at bestå køreprøven til AI.

AI ER OGSÅ FOR DIG

AI Portalen er skabt for at formidle AI konstruktivt og kritisk for alle.

BLIV MEDLEM - STØT UAFHÆNGIG JOURNALISTIK

Som medlem får du adgang til hele magasinet, artikler, events og fællesskab omkring AI Portalen.

Din støtte gør uafhængig journalistik om AI mulig.

AIPortalen.

Det seriøse medie om AI

www.ai-portalen.dk