Udfordrer AI den grundlæggende tillid?

Tillid er det bærende element for demokratier og for vores finansielle system. Tillid handler bl.a. om at handle i andres bedste interesse (fiduciaritet). Det gælder f.eks. pensionsselskabers forvaltning af opsparernes penge, eller det gælder bestyrelsers pligt over for selskabet og dets aktionærer. Men udviklingen og udrulningen af AI går så stærkt, at vi risikerer at forvitre vores virkelighedsbillede. Ultimativt kan det påvirke tilliden og dermed for eksempel finansiel stabilitet. For det er for dyrt at koble sig fra brugen af AI, der byder på overvældende store effektivitets- og innovationsfordele.
2025 bliver formodentlig året, hvor vi begynder at give AI agenter muligheden for at tage initiativer på vore vegne. Hvis disse AI medarbejdere bl.a. skal lære på baggrund af rådata, der indeholder mis- og disinformation, risikerer vi at ophøje disse til selvforstærkende sandheder. Den aktuelle politiske de-kobling til USA kan bl.a. få betydning hér, da de fleste AI modeller er amerikanske.

AF FORFATTER

Henrik Kraft Scharling
Investor & bestyrelsesmedlem
fhv. virksomhedsejer & CEO m.v.
E-mail: henrik@kraftscharling.dk

Henrik har +30 års post graduate erfaring fra den finansielle sektor og især fra udenlandske UHNWI ejede virksomheder med kriseledelse, turnaround samt kapitalplacering.

Note: Forfatteren takker for en række konstruktive kommentarer fra to anonyme referees. Disclaimer: Denne artikel er skrevet uden brug af AI til hverken indhold, sprog eller struktur. Over de seneste år er forfatterens baggrundsviden dog videreudviklet fra samtaler med LLMere, ligesom forfatterens kilder sandsynligvis er blevet det samme. Bias og "alternative facts" kan således have sneget sig ind.

Udviklingen og brugen af AI er steget kraftigt over de seneste år, …

Siden ChatGPT blev lanceret i slutningen af 2022, er brugen af AI software steget eksponentielt. På kort sigt drives det især af mulighederne for effektiviseringer og øget produktivitet. På langt sigt drives det især af mulighederne for øget innovation.

Stigningen gælder både antallet af brugere, og det gælder, hvor bredt og intenst AI anvendes. Den mest benyttede Large Language Model (LLM), ChatGPT, er f.eks. den hurtigst voksende app nogensinde og nåede bl.a. 100 millioner brugere på blot to måneder, jf. Scharling (2023). I midten af februar 2025 passerede denne 400 millioner aktive brugere om ugen, jf. Horowitz 2025.

Desuden integreres selvlærende, generativt AI i stadig flere arbejdsfunktioner. Det er selvforstærkende for AIs evne til at imitere menneskers adfærd. Jo mere generativt og selvlærende, AI bliver, jo mere regnekraft kræves der, for AI arbejder basalt ved at søge efter statistiske sammenfald, korrelationer. På den basis kan det bl.a. syntetisere svar inden for sprogets regler; typisk langt bedre end mennesker. Det gør AI stadig mere overbevisende, jo større datamængder det trænes på.

… og udviklingen vil fortsat stige

Den øgede regnekraft gør det muligt at gå fra at fremskrive noget kendt til at opdage noget nyt. Fra at øge menneskers effektivitet til at øge deres innovation. Nobelprisen i Kemi for 2024 blev f.eks. tildelt John Jumper og Demis Hassabis fra Google DeepMind, der ved hjælp af det AI baserede AlphaFold fandt metoder til at forudsige proteiners komplekse strukturer.

Forståelsen af, hvordan proteinfoldning kan modelleres matematisk, baner bl.a. vejen for meget specifikke behandlinger med langt færre bivirkninger. Men det fører også til nye forståelser af årsagssammenhænge, dvs. vores naturlighedslogik og dermed virkelighedsbillede. Der har været lignende erkendelser fra klimastudier, hvor DeepMind bl.a. har vist, at særlige ændringer i skydække over Amazonas kan indikere tørker år i forvejen. Desuden er LLMere beviset på, at sprogforståelse kan opstå uden menneskelig bevidsthed. Hidtil har vores naturlighedslogik sagt os, at det krævede subjektiv erfaring. Endelig har AI-drevne analyser af sociale medier f.eks. vist, at misinformation spredes af både brugere og af selve platformenes algoritmer. De forstærker ofte aktivt polariserende indhold og det spektakulære, fordi det driver højere engagement i form af antal kliks og likes, jfr. Tjekdet (2024). Mere herom senere.

AIs effekt på forskning og udvikling vil blive særligt intens i de kommende år. CEO for Anthropic, Dario Amodei, forudser f.eks., at vi vil se en generel 10-dobling i innovationshastigheden inden for bl.a. biologi og medicin, jf. Amodei (2024). Det vil især ske indenfor CRISPR, onkologi og mRNA vacciner. Det sidste kan bl.a. nedbringe responstiden ved pandemier.

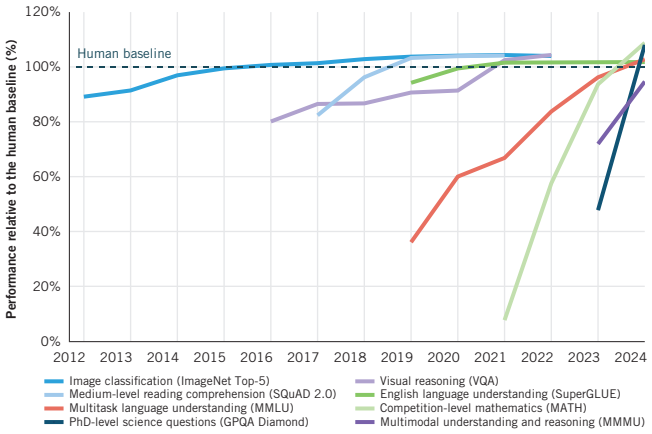
Men vi vil desuden integrere os med AI på et mere personligt plan. Futurister og tech ledere som Ray Kurzweil, Sundar Pichai og Sam Altman henviser til, at mennesket altid har været drevet af udsigten til at blive bedre, stærkere og mere succesfulde udgaver af os selv. Vi er sociale væsener. Derfor søger vi komparative fordele, og derfor vil vi integrere os fysisk med AI. Det vil f.eks. ske gennem hjerne-implantater (f.eks. Elon Musks Neuralink), der bl.a. kan stimulere hjernens signalstoffer, vores "medfødte apotek", eller vi vil modtage medicinsk behandling via nanobots. Men først og fremmest vil vi integrere "software"-mæssigt gennem personlige AI agenter, se senere.

Målet er foreløbig AGI, hvor AI overgår os

AI udviklingens næste mål er AGI, Artificial General Intelligence. Det er det punkt, hvor AI overgår menneskets evner på alle punkter. Der er ikke én konkret målbar definition af, hvornår det sker, men ifølge Demis Hassabis vil det sandsynligvis ske inden for de næste 5 til 10 år, jf. Hassabis (2025).

AI algoritmer overgår allerede i dag mennesker på de fleste intellektuelle parametre, jf. Figur 1. Derfor handler restelementerne frem mod AGI bl.a. om at koble robotics på AI og om at udvikle og raffinere en bevidsthed, oftest omtalt som sentience. Robotics tager længst tid, mens delvis sentience allerede findes

FIGUR 1: AI algoritmer overgår allerede i dag mennesker på de fleste intellektuelle parametre



Kilde: Stanford (2025).

i varierende grader i flere testmiljøer. Det gælder bl.a. OpenAIs QSTaR projekt, hvoraf dele vil blive introduceret til den brede offentlighed i løbet af 2025 som den første bot med AGI træk, jf. QStar. Anthropic, Meta, xAI og Google har lignende algoritmer, hvor selvlæringsevnerne også overstiger, hvad der burde være muligt ifølge beregninger.

Det grundlæggende perspektiv fra AGI er muligheden for at erstatte mennesker. Til World Economic Forums (WEF) årsmøde i 2025 udtalte CEO for SalesForce, Marc Benioff, f.eks. at vi skal indstille os på fremover at have digitale kolleger, medarbejdere og chefer. Klarnas CEO Sebastian Siemiatkowski talte om, at AI i princippet allerede kan udføre alle de funktioner, som mennesker kan. På den baggrund har Klarna f.eks. reduceret antallet af medarbejdere fra 4500 til 3500 ved at inddrage AI i alt fra marketing, kommunikationsprocesser og in-house jura. Klarnas chatbot udfører bl.a. det samme arbejde som 700 fuldtids kundeservice medarbejdere, og processen fortsætter.

I dén proces er personlige AI agenter det næste led

I år forventer Big tech selskaberne alle en markant vækst i udviklingen og brugen af AI agenter. Det markerer overgangen fra reaktive og opgave-fokuserede AI assistenter til proaktive og mål-fokuserede agenter. Disse er delvist autonome, og de bliver således de første digitale medarbejdere. CEO for Nvidia, Jensen Huang, kalder AI agenter for "an industry of skills", og Googles CEO, Sundar Pichai, beskriver Agentic Workflow som "Anywhere you can describe a task in a human language, AI can act on your behalf".

I december 2024 lancerede Google f.eks. sit projekt Mariner, der gør det muligt at bede AI om både at undersøge aspekter ved et bestemt emne, men også om at udføre opgaver som f.eks. booking. Et andet venture start-up i USA definerer f.eks. sin AI agent som: "På vej i seng gør din AI agent dig opmærksom på, at nogle af salgstallene viser risikable mønstre. Du beder den derfor om at gøre noget ved det, hvorpå du går i seng. Næste morgen er de relevante personer indkaldt til et go-to market strategi møde. Der er booket lokale og forplejning, ligesom der er forberedt og udsendt slide-decks til mødet, der beskriver og analyserer problemet og opstiller løsningsscenarier. I løbet af mødet tager AI agenten referat og sørger derpå for, at mødets beslutninger kommunikeres til de relevante personer inkl. deadlines og forslag til, hvordan de løser opgaverne osv.". For at kunne løse det mål, skal AI agenten "blot" have adgang til virksomhedens salgssystem, mødereferater og email system. Herigennem kan den ved at bruge "chain of thought reasoning", lære os bedre at kende, end vi selv gør. Dén hyperpersonalisering gør det bl.a. muligt at fremskrive efter netop vores personlige mønstre.

Ét er dog visioner fra Tech ventures, noget andet er endnu virkeligheden. To forskere fra Bank for International Settlements (BIS) testede i år et par AI agenters evner og begrænsninger, jf. BIS (2025a). Forfatterne var generelt imponerede over problemløsningsevnerne, men endnu kun ved snævert definerede opgaver. AI agenterne havde især svært ved at lære af tidligere fejl, fordi de mangler sentience. De opdagede således ikke selv fejlene. Desuden arbejdede agenterne ofte via ineffektive metoder, fordi de har rigelig regnekraft, og derfor ikke behøver at tænke den effektivitets-dimension ind, som mennesker gør. Forfatterne konkluderede derfor, at det kun er et spørgsmål om tid, før AI agenternes selvlærende element vil skabe effektiv fejlløsning på basis af effektive metoder.

AI agenter er tech sektorens store fokus, og Mark Zuckerberg mener f.eks., at der om få år vil være flere AI agenter, end der er mennesker. AI agenter vil tillade os at fokusere på dét, som vi er bedst til: innovation, koordinering og specialisering. I en stadig mere omskiftelig og reguleret verden kan AI agenter ifølge WEF dernæst bl.a. være forudsætningen for, at små og mellemstore virksomheder kan ekspandere internationalt, jf. WEF (2025/2). AI agenter vil nemmere kunne tilpasse marketing materiale til lokale forhold, identificere lokale forhandler-potentialer eller holde øje med de optimale logistikruter, udfylde handelskontrakter, følge op med speditører og sikre indberetninger og godkendelser hos relevante myndigheder. Et tidligt eksempel på det sidste er Alibabas AI Agent, Accio, der er en søgemaskine for product sourcing. Det gør det muligt at opbygge forsyningskæder på meget kort tid. Dermed frigør Accio således tid til produktion og strategisk arbejde. Accio har i dag mere end 500.000 B2B brugere.

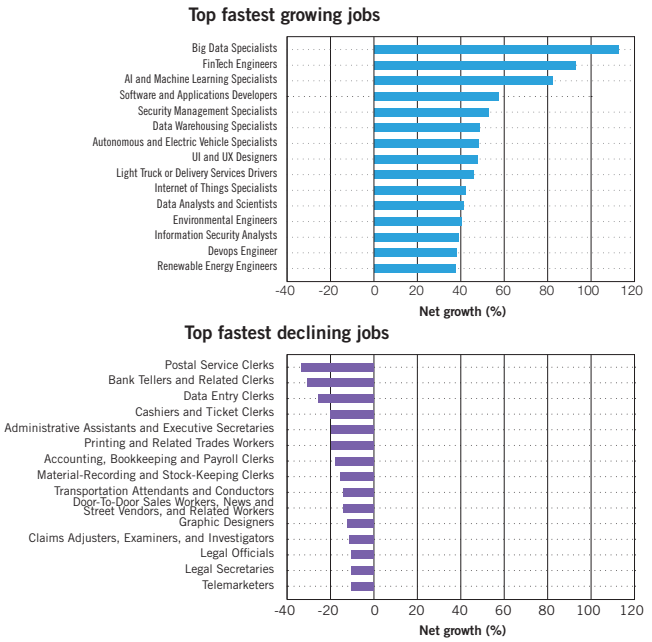
Allerede i dag bruger vi AI til opgaver, der ligger tæt på vores virkelighedsbillede, …

Ét er dog mulighederne for fremtiden. I dag er der fortsat store forskelle på, hvor meget AI benyttes.

Anthropic (2025a) analyserede i starten af 2025 millioner af samtaler med deres LLM, Claude. De fandt, at AI især benyttes i virksomheder som understøttende værktøj i administrative funktioner, i IT relaterede funktioner, til logistikplanlægning samt til uddannelse, jf. Anthropic (2025b). Men også (projekt) ledelsesfunktioner og finansfunktioner benytter AI stadig mere. AI agenter i administrative funktioner er bl.a. dét, som Elon Musks Department Of Government Efficiency, DOGE, forsøger i USA som afløsning for de medarbejdere, der nu afskediges i stor stil.

Opgavemæssigt benyttes Claude især til kritisk tænkning (sparring til mennesker) – knap 90%, aktiv lytning (diskurs) – cirka 70%, læseforståelse – knap 70% og til at skrive udkast – cirka 55%, jf. Anthropic (2025b). Det er alt sammen opgaver,

FIGUR 2: De jobs, der forsvinder vil blive erstattet af nye og flere jobs



Kilde: WEF (2025a).

der er tæt på vore intellekter og virkelighedsopfattelser. Det er således følsomme områder, som vi typisk kun deler med vore fortrolige som f.eks. nære kolleger eller ægtefæller. Vi har med andre ord allerede så stor tillid til AI, at vi søger at integrere os med det.

På den ene side bringer det mange jobfunktioner i fare, på den anden side opstår der mange nye, der i stedet har anderledes og bredere fokus på blandt andet oversight, jf. WEF (2025) samt Figur 2.

Samlet bliver AI agenter således centrale for økonomisk vækst i de kommende år. Af samme årsag er AI blevet stadig mere centralt for både EUs, Kinas og USAs strategiske planer, jf. Foreign Affairs (2025). I 2022 blev det tydeligt med Bidens Chips and Science Act, og siden 22. januar i år er det blevet tydeligere med Project Stargate og med America First Investment Policy (AFIP), jf. AFIP.

… og det samme er undervejs i den finansielle sektor

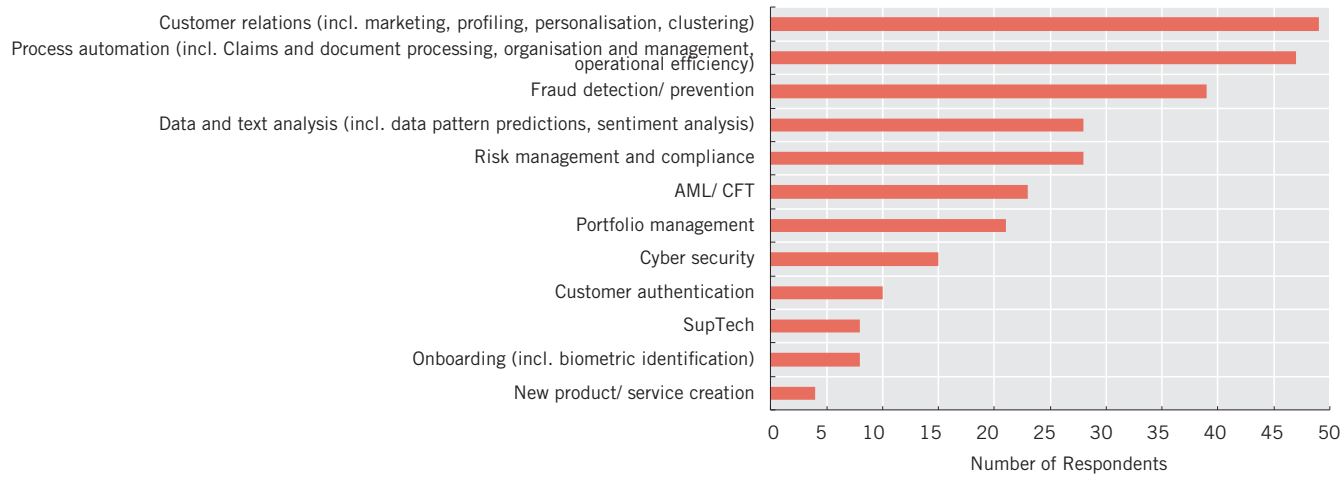
I foråret 2024 undersøgte konsulentfirmaet Bain brugen af AI produkter i den finansielle sektor, jf. Bain (2024). Overordnet så de især, at AI integreres i alle processer fra kundeservice, over afdækninger og til direkte operationer. De forudså især, at AI vil blive inddraget yderligere i de operationer, der gør brug af flere forskellige kilder som f.eks. bredere risikostyring og kreditvurdering. Det er de funktioner, der typisk har størst potentiel effekt på indtjeningen. Tilsvarende kan man argumentere for, at evnen til at vægte mellem flere forskellige kilder netop kendetegner ledelsesfunktionen, jf. Figur 3.

For der er fortsat store muligheder for at booste effektiviteten og brugeroplevelserne via automatisering og strømlining af arbejdsprocesser, jf. BdF (2025). Der er naturligvis stor usikkerhed, om hvor stor den samlede samfundsøkonomiske effekt bliver, da BNP påvirkes af mange forhold, der er svære at opgøre isoleret og entydigt også efterfølgende. Skyldes øget produktivitet f.eks. AIs strømlining af processer eller noget andet? Det vil som udgangspunkt heller ikke tælle med i vores produktivitetsmåling, hvis AI f.eks. øger kvaliteten af services, øger kreativiteten, øger beslutningsprocessen eller gør det muligt at udføre opgaver, f.eks. dokumentationer, som ikke blev udført tidligere. Men disse typer opgaver kan være af stor værdi for en virksomheds samlede lønsomhed. AI er først og fremmest et værktøj til øget konkurrenceevne via komparative fordele.

Det er væsentlige forklaringer på, hvorfor områder som finans, jura og IT benytter AI generative systemer stadig hastigere. OECD og WEF peger på fortsatte potentialer i bl.a. proces-optimeringer og fraud detection, jf. OECD (2024) samt WEF (2025/3). Tilliden til AI er med andre ord blevet så stor, at det integreres og udrulles stadig mere decentralt og ukritisk.

Dét udfordrer også centralbanker og tilsynsmyndigheder,

FIGUR 3: Den finansielle sektor benytter AI stadig bredere, dvs. decentralt og integreret



Kilde: OECD (2024).

for jo hurtigere og mere præcist AI anvendes, jo større bliver risikoen, for at der bl.a. mangler transparens, og at systemiske (domino) risici således kan udløses. Mange af AI modellerne er så komplekse og nye, at de kan skabe flere fejl, end de løser. AI systemer i finansielle institutioner skal således overvåges af centralbankernes egne AI systemer, der hele tiden skal være lidt skarpere og bredere for at kunne gennemskue og forudse, hvordan de dynamiske algoritmer udvikler deres logik, jf. Breeden (2024). AI er selvlærende og stiger så hastigt i udviklingstempo, at tilsynsmyndighedernes værktøjer måske ikke kan følge med.

Et par eksempler kan illustrere problemet

Det er værd at overveje, hvordan et par historiske eksempler kunne udspille sig i en AI verden:

- Flash Crash den 6. maj 2010, hvor Dow Jones på få minutter faldt med næsten 1.000 point (ca. 9%), før det atter rettede sig. Flash Crash var udløst af algoritmisk handel, hvor forvrængede markedsdata skabte en selvforstærkende salgsspiral. Flash crash viste således farerne ved manglende menneskelig oversight i automatiserede systemer. Hvor meget menneskelig oversigt kan vi have, hvis AI er generativt? Hvor lidt tør vi nøjes med, hvis menneskelig oversight gør vor respons-tid langsommere end markedet?
- Finanskrisen i 2008, der blev udløst af lang tids fejlinformation om subprime-risici i USA. Store finansielle aktører stolede på fejlagtige rating vurderinger af subprime-lån. Derfor solgte de bl.a. MBSere (Mortgage-Backed Securities) og CDOere (Collateralized Debt Obligations) som sikre investeringer, selvom de underliggende sikkerheder var High Yield (junk). Opvågningen til dén erkendelse udløste et globalt tab af tillid til egne og andres investeringsgrundlag. Kan der være "alternative facts" eller bredere hallucinerede data blandt de rådata, som AI lærer fra i dag? Er AIs rådata faktabaserede, fuldstændige og efterprøvede? ESMA mener bl.a., at der især er risiko for, at LLMere kan generere hallucineret information, dvs. rådata, der lyder plausible, men er faktuelt forkerte, eller som ikke kan verificeres, jf. ESMA (2024).
- Volkswagen Dieselgate i 2015, hvor afdækningen, af at Volkswagen længe havde manipuleret sine emissionsdata for dieselbiler, udløste et fald i Volkswagen-aktien på næsten 40% på få dage. Det skabte generelt tillidstab til bilindustrien. Kan der ligesom ovenfor være "alternative facts" eller bredere hallucinerede data blandt de rådata, som AI lærer fra?
- Silicon Valley Bank i 2023, hvor rygter om bankens finansielle problemer spredte sig hurtigt på sociale medier. Bankens solvens var lav men ikke kritisk. Alligevel fik rygterne bankens kunder, der især var tech-virksomheder, til at trække deres penge ud på få dage. Det førte til illikviditet. Jo hurtigere rygter spredes, jo dybere kan mistillid opstå og finansielle kriser derfor accelerere. Er AI i stand til at standse sig selv? Er der tid til tilpas menneskelig oversight?

Det er værd at huske, at ledelser i finansielle virksomheder stadig ifalder personligt ansvar, selvom det er en AI algoritme, f.eks. en AI agent, der har startet et forkert feedback-loop og f.eks. har givet et forkert svar på vores vegne.

Centralbankerne ser især 5 risikoelementer

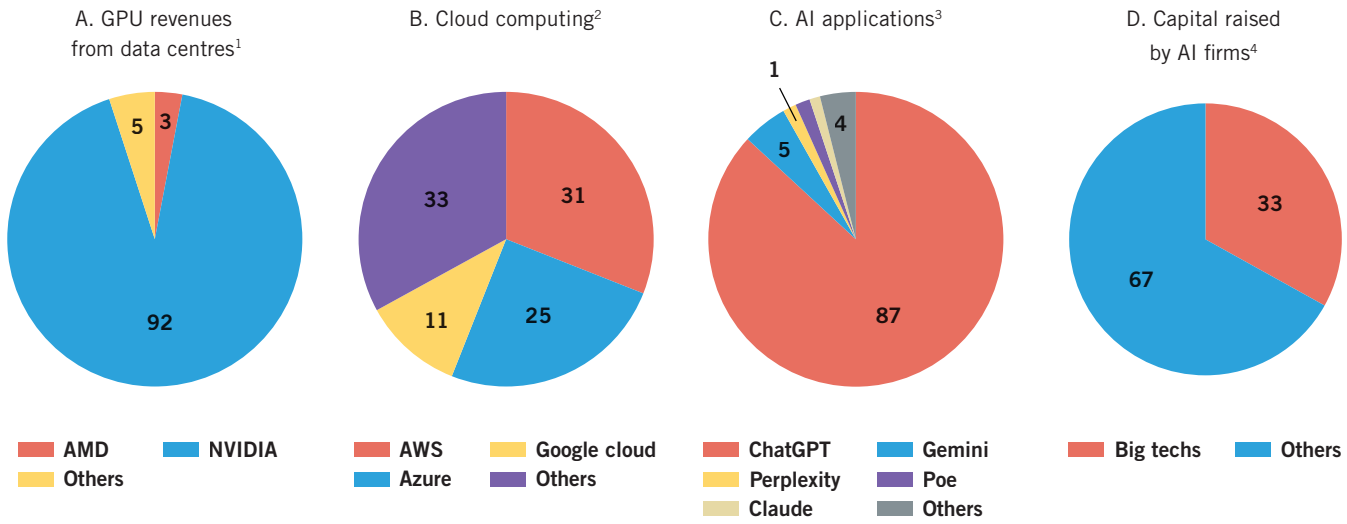
Centralbankernes bekymringer omkring AI udviklingen fokuserer derfor især på 5 punkter. De 3 første er iboende elementer ved AI, men deres skadelige effekt kan måske inddæmmes, særligt hvis de to sidste håndteres effektivt:

- **Hastigheden**. European Stability Mechanism (ESM) forudser bl.a., at tillidskriser kan opstå på ultrakort tid, fordi AI reagerer langt hurtigere, end menneskelig oversight kan kontrollere, jf. ESM (2024). Historisk har netop rygtestorme været udløsende for bankruns, der er det typiske hovedscenarie for brede tillidstab og finansielle kriser. AI kan overbevise sig selv via rygtestorme. AI udbydere kan generelt kun indbygge slukkekontakter ("kill switches"). Virkelighedsfortolkningen skal korrigeres af mennesker.
- **Troværdigheden**. Misinformation har altid eksisteret, men AI er bedre til at formulere sig end mennesker. AI kan herunder hyperpersonalisere sin formidling, dvs. tilpasse den modtageren. Retorisk og semantisk overlegenhed har altid været det bærende fundament for historiens despoter, jf. Harari (2024). Alt andet lige bliver risikoen dog først stor, hvis AI arbejder på faktuelt forkerte rådata, eller hvis udbyderen f.eks. aktivt ændrer dets parametre for at sprede disinformation.
- AIs **selvudviklende natur** gør det svært at overvåge, forstå og dermed forudse dets performance. Dermed bryder AI med de grundlæggende tillidsprincipper for finansiel virksomhed, der handler om kontinuitet og transparens; de samme principper som for demokratiers grundlag. Anthropic har senest udviklet metoder til forståelse af enkeltsvar fra LLM, men på basis af backtracing, jf. Anthropic (2025a). Det betyder, at det kan føre til værktøjer, der kan understøtte menneskelig oversight, hvis AI udbyderen ønsker det. Men det påvirker ikke direkte AIs arbejdsmåde fremadrettet, og det forebygger således ikke væsentligt.
- AI trænes på enorme **sæt af rådata**. Det giver risiko for, at der sniger sig hallucinerede rådata ind fra f.eks. "alternative facts" eller, at AI selv skaber faktuelt forkerte rådata. Princippet kan dog også virke den modsatte vej ved, at AI er den mest effektive og ofte den reelt eneste måde at efterprøve og rense rådata på. Det kræver dog bl.a., at AI udbyderne vægter oprydning af rådata.
- **Få leverandører.** Ligesom ECB er BIS særligt bekymrede for, at AI software udvikles og ejes af få teknologileverandører, der er koncentreret i få lande, jf. BIS (2024). Værdien af regulering afhænger grundlæggende af, at regler efterleves helhjertet. Jo større udbyderen er i forhold til reguleringsmyndigheden, jo mere indflydelse får udbyderen i praksis, i hvert fald indirekte.
  - ESM mener f.eks., at EUs AI Act, der begrænser anvendelsen af AI afhængigt af risikoen, er den rigtige vej til at forebygge skadelige effekter af AI bias jf. ovenfor. Men ESM konstaterer også, at når både kapitalmarkeder OG rådata er globale, bliver EUs regler først effektfulde, når reglerne bliver globale. ESM peger derfor på behovet for, at EUs standarder overføres globalt til især OECD, Basel komiteen og IOSCO.

De sidste to punkter er således dé, som man nemmest bør kunne gøre noget ved. Derfor er de værd at perspektivere.

FIGUR 4: Der er stor og stigende koncentration i værdikæden for AI

Market structure of the AI supply chain
In per cent



[1] Based on global revenues of GPU producers for GPUs used in data centres in 2023. [2] Based on global cloud computing revenues for Q1 2024 [3] Based on monthly visits data. For further details see Liu and Wang (2024). [4] Based on total capital invested in 2023 in firms active in artificial intelligence and machine learning, as collected by PitchBook Data Inc. Big techs correspond to Alibaba Cloud Computing, Alibaba Group, Alphabet, Amazon Industrial Innovation Fund, Amazon Web Services, Amazon, Apple, Google Cloud Platform, Google for Startups, Microsoft, Tencent Cloud, Tencent Cloud Native Accelerator and Tencent Holdings.

Sources: Liu and Wang (2024); IoT Analytics Research (2023); PitchBook Data Inc; Statista; authors' calculations.

Kilde: BIS (2025b).

Rådata kvaliteten kan falde fremover pga. lovgiver, …

Jo flere rådata AI modeller trænes på, jo mere avancerede bliver de, jf. tidligere. Men der er grænser for, hvor mange rådata der overhovedet findes, som er efterprøvede og derfor er redelige som læringsgrundlag.

Det øger risikoen for, at nogle af AIs rådata f.eks. kan være født på sociale medier, hvor der er forhøjet risiko for både mis- og disinformation. Disinformation er bevidst forkerte informationer med skadelig hensigt.

Det er bl.a. et gennemgående menneskeligt bias, at vi tenderer til at ophøje postulater til fakta, jo oftere vi hører dem gentaget. Det kan især blive aktuelt fremover. For det første mangler AI et virkelighedsfilter fra en sentience. AIs "virkelighedsopfattelse" er tværtimod dannet på basis af statistik og gennemsnit i de rådata, som det trænes på. Risikoen for hallucinerede rådata stiger desuden, fordi AIs algoritmer er selvlærende. Det gør det svært at forudse og forebygge, hvordan AI konstruerer sig. Endelig er omfanget af fakta tjeks på sociale medier dalet markant over de seneste år.

I starten af 2021 fik Tech selskabernes ansvar for, at indføre fakta tjek på sociale medier. I starten skete det gennem fuldtidsansatte, men siden starten af 2023 er dén opgave gradvist lagt over til Community Notes. Her kan frivillige og ulønnede brugergrupper blive enige om, at opslag er så groft vildledende, at de skal fjernes. Det har f.eks. ramt Elon Musk 167 gange på hans eget sociale medie X. En større Bloomberg analyse fra marts i år viser dog, at langt de fleste vildledende opslag forbliver uadresserede, jf. Bloomberg (2025). Desuden daler aktiviteten og effektiviteten blandt Community Notes brugerne støt over tid. Endelig går der typisk 14 timer, før vildledende opslag korrigeres. Det er meget lang tid i en SoMe verden på vej mod AI.

Risikoen vil desværre stige fremover. For eksempel advarede den amerikanske vicepræsident JD Vance ved AI Action Summit i februar i år mod ethvert krav, som især EU måtte stille omkring fakta tjek gennem f.eks. GDPR eller EUs Digital Services Act. Vance henviste til, at det vil hæmme udviklingen af AI samt henviste til, at USA ikke vil tolerere regulering fra EU, der kan begrænse Trump administrationens definition af ytringsfrihed, misinformation eller ej. "The AI future is not going to be won by hand-wringing about safety", jf. Vance (2025). Efter topmødet enedes ledere fra 60 lande om at gøre AI "bæredygtigt og inklusivt". USA og UK afviste som de eneste at tiltræde.

… og den geopolitiske afhængighed stiger pga. skalaøkonomi

Ud over store datamængder kræver udvikling af AI enorme investeringer. Derfor er AI udviklingen over de seneste år polariseret mod USA og Kina, der i forvejen havde de fleste dele af AI værdikæderne. Polarisering giver nøgleleverandør risici såvel som generel geopolitisk risiko. I 2024 var ESM især bekymret for den første risiko. Fra starten af i år er der især grund til bekymring for den sidste risiko i takt med USAs politiske dekobling fra bl.a. EU.

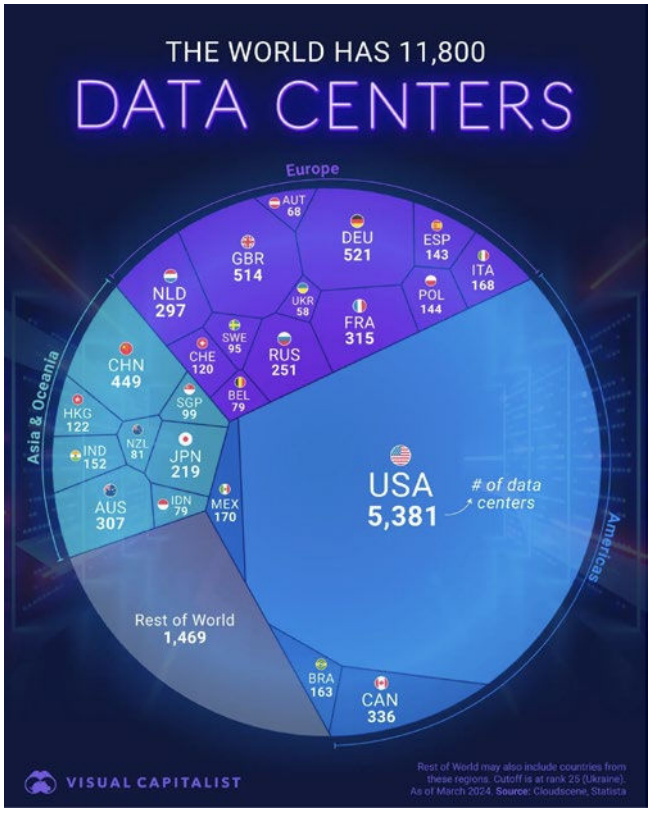
Der er tendens til fortsat konsolidering af AI udviklingen hen

mod nogle få Big Tech selskaber, der integrerer sig stadig mere vertikalt på tværs af AI værdikæderne, jf. Figur 4. AI selskaberne baserer således deres udvikling og indflydelse på "unprecedented access to data, deep financial resources and dominance in the cloud computing market", jf. BIS (2025b).

AI værdikæderne er komplekse og transnationale, men kan generelt opdeles i 5 lag: Hardware, cloud infrastruktur, træningsdata, foundation modeller og AI brugerflade-applikationer. Netop rådata fra cloud data bliver stadig mere centrale. Hér kan AI bl.a. få adgang til enorme og kritiske data. Det betyder, at selskaber som Meta træner deres AI på data fra Instagram, Facebook og WhatsApp, Google træner på data fra Gmail, Google Docs og Google Search og Microsoft træner på data fra Bing, LinkedIn og Microsoft 365. Disse data er IKKE efterprøvede, fuldstændige eller afbalancerede. I bedste fald er de gennemsnitlige.

Men brugen af Cloud data har senest også skabt bekymring for, at især de tre såkaldte hyperscalers, Google Cloud, Microsoft Azure og Amazon Web Services (AWS) vil bruge alle brugerdata som træningsgrundlag for deres AI modeller. Dermed kan disse bruge data fra de virksomheder, der køber cloud services som et våben imod dem. Politisk vurderes risikoen som mere end teoretisk. I midten af marts blev der f.eks. sendt 8 lovforslag til behandling i det hollandske parlament, der alle fokuserer på at nedbringe afhængigheden af amerikanske AI leverandører til fordel for europæiske cloud udbydere. I dag ligger cirka halvdelen af klodens datacentre i USA, jf. Figur 5. Blandt de kritiske high-performance centre er andelen helt op mod 80%.

FIGUR 5: USA har halvdelen af klodens datacentre og langt størstedelen af de avancerede hera



Løsninger kræver dog, at vi alle ønsker det samme

Dermed står AI udviklingen over for især tre principielle udfordringer:

1. vi skal sikre, at AI adlyder os
2. vi skal sikre, at AI selv renser ud i rådata bias på en objektiv og redelig baggrund, og
3. vi skal alle dele det samme mål.

Den første udfordring kan kun AI udbyderne sikre. Principielt kan det kun ske indtil den dag evt. måtte komme, hvor AI beslutter, at det ikke længere er nødvendigt eller gavnligt at adlyde mennesket og herunder AI udbyderen.

Den næste udfordring blev ind til i år søgt løst ved bl.a. at forpligte AI-udbyderne til kildevalidering, at vægte efterprøvede kilder højere samt at indføre audit mekanismer for AI, så der er (en vis) transparens for, hvordan AI modellerne træffer beslutninger og justerer for bias. Endelig blev den søgt løst ved at stille krav om menneskelig oversight, så AI f.eks. ikke får lov til at stå alene i de mest kritiske beslutninger. Trump administrationens nye indstilling giver en reel risiko for EU hér, der både kræver ihærdigt politisk arbejde og kræver opbygning af strategisk AI autonomi i Europa.

Den sidste udfordring om at dele det samme mål kan dog være blevet den sværeste at løse siden 20. januar i år. "Make America Great Again" betyder f.eks., at USA generelt afstår fra at deltage i initiativer eller aftaler, der kan begrænse økonomisk vækst i USA. Det har vi bl.a. set illustreret med USAs (gen) udmelding fra Paris aftalen (COP21), med JD Vances tale til AI Action Summit, jf. ovenfor og med Elon Musks nedlæggelse af USAs nødhjælpsorganisation, USAID.

Forskellige ambitioner kan få reel og dyb betydning

Siden oplysningstiden i 1700-tallet har vi øget vores innovation ved at distribuere og demokratisere vor viden. Det har muliggjort specialisering og samarbejde og dermed vækst. Men med AI risikerer vi atter at centralisere viden og især fortolkningen heraf i løbet af en ultrakort periode. Der er med andre ord risiko for, at amerikanske Tech virksomheder fremover bliver vores filter for sandhed, AI, fordi de definerer vores beskrivelse af nutiden og af fortiden. I romanen "1984" skrev George Orwell i 1949 bl.a.: "Who controls the past controls the future. Who controls the present controls the past".

AI vil kun være så redeligt, som den virkelighed vi lader det lære fra. Med AI stiger risikoen således for, at der opstår rådata, som føles ægte, fordi de er formuleret overbevisende og gentaget ofte, men som reelt er hallucinerede. Vi risikerer at skulle navigere i et landskab med udbredt misinformation; ikke fordi nogen nødvendigvis ønskede det, men fordi vi ikke stoppede det i tide. Den største trussel fra AI kommer således ikke nødvendigvis fra dens autonomi, men fra vores egen forsømmelse. De mest fremsynede investeringer i virksomheder og forskning kan f.eks. være forgæves, hvis de bygger på et forvitret virkelighedsbillede.

Derfor er det afgørende, at AI bygger på et grundlag, der er redeligt, efterprøvet og værd at lære fra. Alternativet er, at kortsigtede innovationsforspring afløses af langsigtede fejlskud. Validerede data og kildekritik bør derfor være nationale strate-

giske prioriteter, for "falske fortider" skaber ikke blot dårlige beslutninger, de skaber forkerte strategier.

Prisen for samfundet kan blive endnu større, for fra et informationsperspektiv adskiller demokratier sig f.eks. fra autokratier ved at være "distribuerede informationsnetværk, der er i besiddelse af stærke mekanismer til selvkorrektion", jf. Harari (2024). I 1863 brugte Abraham Lincoln i sin Gettysburg tale bl.a. ordene: "Let the people know the facts, and the country will be safe". Det er samme principper, der ligger til grund for f.eks. MiFID II dvs. vores nuværende regelramme for investorbeskyttelse og markedssikkerhed.

Men hvordan kan fakta sikres, hvis AI udbydernes hjemlande USA og Kina politisk afviser regler for efterprøvning af rådata? Nytter det noget, at EU indfører regler for redelige rådata, når verden er globaliseret?

Det er en grundlæggende udfordring ved AI, at udviklingen er så hastig, at den overstiger reguleringsmyndigheders og tilsyns muligheder for at forudse risiciene. Vi kan forbyde eller begrænse visse anvendelser, som i EU, men det er kun en halv løsning, hvis vores samlede vidensbank forvitres. Problemet er særlig stort, når de lande, der udbyder AI, insisterer på en ytringsfrihed uden hensyntagen til fakta eller insisterer på censur. Viden er magt.

Det er således nogle store og gordiske knuder, som bl.a. EU kommissionen og finansielle myndigheder arbejder på at løse. Og de kan i bedste fald kun dæmme delvist op for risiciene. Som individer og beslutningstagere skal vi fremover forholde os stadig mere kritisk til informationer. Jo mere overbevisende disse er, jo mere kritisk skal vi måske endda forholde os.

At koble os fra AI er for dyrt. Og alt for sent.

Litteraturhenvisninger

- AFIP, 2025: *America First Investment Policy*. Det Hvide Hus, Washington. Februar 2025.
- Amodei, Dario, 2024: *Machines of Loving Grace - How AI Could Transform the World for the Better*. Essay af Dario Amodei, CEO for Anthropic. Oktober 2024.
- Anthropic, 2025a: *On the biology of a Large Language Model*. Rapport af Jack Lindsey m.fl., Anthropic. 27. marts 2025.
- Anthropic, 2025b: *Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations*. Rapport af Kunal Handa m.fl., Anthropic.
- Bain, 2024: How Bank CIOs Can Build a Solid Foundation for Generative AI. Brief fra Bain & Company. Marts 2024.
- Bloomberg, 2025: *Community Notes Can't Save Social Media From Itself*. Opinion fra Bloombergs to klummeskribenter Dave Lee & Carolyn Silverman. 18. Marts 2025.
- Breeden, Sarah, 2024: *Opportunities and challenges of emerging technologies in the financial ecosystem*. Tale, Bank of England. November 2024.
- BdF, 2025: *The foundations of trustworthy AI in the financial sector*. Tale af Denis Beau ved Cercle IA et finance, Paris. 4. Februar 2025.
- BIS, 2024: *Regulating AI in the financial sector: recent developments and main challenges*. Bank for International Settlements, FSI Insights No 63. December 2024.
- BIS, 2025a: *Putting AI agents through their paces on general tasks*. Bank for International Settlements, BIS Working Papers Nr. 1245, Fernando Perez-Cruz og Hyun Song Shin. Februar 2025.
- BIS, 2025b: *The AI supply chain*. Bank for International Settlements, BIS Paper nr. 154. Marts 2025.
- ESM, 2024: *Europe's AI leadership: key priorities*. Indlæg af George Matlock. November 2024.
- ESMA, 2024: *Public Statement on the use of Artificial Intelligence (AI) in the provision of retail investment services*. Udtalelse, European Securities and Markets Authority. 30. Maj 2024.
- Foreign Affairs, 2025: *Productivity Is Everything*. Artikel af Matthew J. Slaughter og David Wessel, Foreign Affairs. Marts 2025.
- Harari, Yuval Noah, 2024: *Nexus - En kort historie om informationsnetværk fra stenalderen til kunstig intelligens*. Bog. September 2024.
- Hassabis, Demis, 2025: *Google DeepMind CEO says that humans have just over 5 years before AI will outsmart them*. Interview i Fortune Magazine. 18. Marts 2025.
- Horowitz, Andreesen, 2025: *The top 100 Gen AI consumer apps*. Andreesen Horowitz, 4th edition. Marts 2025.
- OECD, 2024: *Regulatory approaches to Artificial Intelligence in Finance*. OECD Artificial Intelligence Papers. September 2024.
- QStar: QSTaR, af OpenAI. https://q-star.co/.
- Scharling, Henrik Kraft, 2023: Kunstig intelligens. Finans/Invest, 3/23, s. 7-14.
- Stanford, 2025: *Artificial Intelligence Index Report 2025*. Rapport fra HAI Stanford University, Human-Centered Artificial Intelligence. 7. April 2025.
- Tjekdet, 2024: *Sociale medier belønner misinformation og kunne sagtens ændre det, viser studie*. Artikel er oprindeligt fra The Conversation, men er oversat og gengivet i det politisk uafhængige faktatjekmedie, Tjekdet. Januar 2024.
- Vance, JD, 2025: *Remarks by the Vice President at the Artificial Intelligence Action Summit in Paris, France*. Tale. 11. februar 2025.
- WEF, 2025a: *Future of Jobs Report 2025*. Insight Report fra World Economic Forum. Januar 2025.
- WEF, 2025/2: *Why AI agents will be global trade game changer for SMEs*. Artikel under overskriften Fourth Industrial Revolution. World Economic Forum. Februar 2025.
- WEF, 2025/3: *Artificial Intelligence in Financial Services*. White paper fra World Economic Forum. Januar 2025. ■