

Digitale visjoner – hinsides instrumentalistiske mistak?

Skjervheims forfatterskap inneholder en kritisk diagnose og advarsel: mot fremskreden objektivisering og instrumentalisering av mennesket.

I en samtale, for eksempel mellom en terapeut og en pasient kan man ikke gjøre den andre til en ting eller et objekt, uten å ødelegge samtalen, tilliten og kanskje også relasjonen. Man kan heller ikke fullstendig objektivisere seg selv, uten å ende i en merkverdig form for fremmedgjøring. Samtalen (med andre og med en selv) forutsetter at det også finnes en subjekt-subjekt relasjon; ikke *bare* subjekt – objekt.

Tilsvarende gjelder i politikken: man kan lage statistiske forklaringer basert på store mengder data, og således objektivere sine medmennesker. Men det man ikke bør gjøre er å erstatte arenaer for politisk og kollektiv meningsutveksling med data, modeller og statistiske analyser: da står vi overfor det Skjervheim kalte kategori-mistak og instrumentalistiske mistak.

‘Ein lever i kvar si verd, der alter er eit faktum for ego i ego si verd’, sier Skjervheim. Man deler ikke lenger en felles verden. Dette skriver Skjervheim inn i en hermenutisk tradisjon av tenkere som ser samfunnet som et kommunikativt fellesskap. Dette fellesskapet trenger felles forestillinger og arenaer for meningsutveksling for å henge sammen.

Dette er et utgangspunktet, som jeg i stor grad deler. Spørsmålet jeg skal stille i dette foredraget innebærer utvidelse av problemet, muligens hinsides horisonten Skjervheim opererte innefor. Jeg stiller spørsmålet om hva som skjer idet vi introduserer en tredje aktør, mellom alter og ego så og si, til fellesskapet. Denne tredje aktøren er en maskin, men ikke en hvilken som helst maskin: det er en maskin som kommuniserer og samhandler, og som innehar menneskelignende egenskaper. Datamaskiner, i sitt mangfold, er gode til å behandle tall, og i den forstand forsterker de potensialet for objektivisering på alle fronter: i forholdet en har til seg selv (for eksempel smarte klokker og sensorer på kroppen), til andre (sosiale medier), til institusjoner og infrastruktur (nye former for samhandling) og til politikk (nye problemstillinger, relasjoner og former for makt oppstår).

De er ikke bare kraftige kalkulatorer; de utgjør et nytt medium, det digitale, som inngår i nær sagt alle relasjoner. Og, de utgjør sentrale metaforer for omfortolkning av relasjoner, ofte (men ikke nødvendigvis) i form av et maskin-sentrert meningsunivers. I resten av dette foredraget fokuserer jeg på hvordan disse menneske-maskin-menneske relasjonene har oppstått (historie), hvordan de har spredt seg i samfunnet, som strategiske instrumenter markeds- og innovasjons- prosjekter (visjoner), og hvordan de i

den senere tiden har inntatt ikke bare kulturen men også de politiske institusjonene og arenaene, inkludert som tekno-ideologi.

Skjervheim var ikke spesielt opptatt av hverken naturvitenskap eller teknologi, bortsett fra når disse beveget seg over på samfunnsvitenskapenes, pedagogikkens eller politikkens områder. Men om vi nå fokuserer på nåtidens mest omdiskuterte teknologi, kunstig intelligens, ville Skjervheim fort dra kjensel på det intellektuelle og historiske utgangspunktet, som var tett sammenvevd med de positive antakelsene som han kritiserte. BILDE 2

Nærmere bestemt sikter jeg til forestillingen, like gammel som moderniteten selv, om et universelt språk, grunnet i en mekanistisk og matematisk objektifiserbar verden. Allerede Leibniz (1700-tallet) forsøkte å skape et formalistisk matematisk og symbolsk språk, der han forsøkte å innarbeide både Gud og metafysikk. Deler av denne ideen ble viderebragt av de logiske positivistene og analytisk filosofi i det 20de århundret. Til forskjell fra Leibniz var de logiske positivistenes eksplisitte ambisjon å rense språket for verdier, metafysikk og religion, men de delte altså forestillingen om et universelt språk.

Parallelt med at positivistene bygget sitt program, ble andre former for program bygget, bokstavelig talt. De første datamaskinene innførte nye dimensjoner til prosjektet om et universelt språk. Briten Alan Turing eksperimenterte med ideen om maskinell prosessering av matematiske symboler. Den underliggende antakelsen for en Turing-maskin (et tanke-konstrukt) var at 'tenkning' egentlig betegner den samme generelle prosessen i maskiner som i mennesker.

Når teorier om neurale nettverk, som i stor grad har inspirert dagens språkmodeller, ble introdusert, var en eksplisitt intensjon å overkomme skillet mellom menneske og maskin. En av grunnleggerne av såkalt symbolsk KI, Marvin Minsky, uttalte hvordan det ikke finnes noen 'grense mellom en teori om menneskelig tenkning og en plan for å lage en intelligent maskin' (Minsky, 1974, 6). Idet Noam Chomsky innvendte at maskiner ikke kan tenke, var tilsvaret fra Minsky at psykologi og læring ikke egentlig er vesensforskjellige fra maskinelle prosesser. De kan derfor behandles som problemer av og for ingeniører.

Kybernetikk er læren om kontroll og kommunikasjon i maskiner og levende vesener, og av mange regnet som KIs forløper. Feltet oppstod i tiden rundt andre verdenskrig, sammen med utviklingen av rakettsystemer og luftforsvar, men også atombomben og atomkraft (tidlige datamaskiner ble brukt for å beregne sprengkraft i hydrogenbomber). På den ene siden hadde man da utviklingen av selvstendig handlende ('autonome') systemer rettet inn mot kommunikasjon og kontroll. På den andre siden en teknologi med nok kraft til å utslette livet på kloden. Til sammen inspirerte disse teknologiene

nær gude-lignende forestillinger om ingeniørers skaperkraft. Teknologiens ustadige forhold til religiøse oppfattelser har vært et tema lenge, som da kybernetikkens grunnlegger, Norbert Wiener skrev boken *God & Golem Inc.* BILDE 3. Wiener skrev boken for å advare mot tendenser som allerede da var til stede. Jeg siterer fra Geoffrey Bowker, *How to Be Universal: Some Cybernetic Strategies*, 1943-70. Bowker gjengir der den britiske kybernetikeren og psykologen Frank George (1965):

‘Det er ingen grunn til å betvile muligheten av å skape et kunstig menneske’;

‘Det som historisk sett har blitt tydelig som menneskets dominerende mål, er dermed replisering av seg selv ved hjelp av teknologiske midler. Formen på dette dominerende målet blir dermed en supermaskin, selvopererende, selvinstruerende og menneskekontrollert, selv om denne sistnevnte prosessen kan reduseres til et minimum...

Dataviteren Joseph Weizenbaum var blant de første til å utforske og utfordre programmet om å gjenskape menneskelig intelligens (70-tallet), et prosjekt Weizenbaum var svært kritisk til. For å bevise det feilaktige i antakelsene det bygde på, skapte han verdens først chat-bot, ELIZA. ELIZA var en etterligning av en samtaleterapeut som kunne svare på enkle spørsmål. Weizenbaum ble imidlertid tatt på sengen, idet det viste seg at folk lot seg forføre av ELIZA. Hans egne kolleger betrodde selv intime detaljer til maskinen, og noen begynte å utvikle personlige relasjoner til den.

Det som særlig overrasket Weizenbaum var den sterke trangen til å projisere menneskelige egenskaper over på ting og maskiner, så-kalt antropomorfisering. For å forklare dette la meg ta et kanskje litt outrert eksempel: Jesus på korset. Om man utgår fra et reint observasjons-språk er det man ser en tre-figur som fremstiller en vanlig mann (i en uvanlig positur). Denne figuren er investert med følelser og forestillinger sterke nok til å sentrere en hel kultur (den vestlige). Noe tilsvarende gjelder for digitale maskiner: de er materielle artefakter som opererer gjennom algoritmer og kode. Samtidig er disse maskinene investert med noen av vår kulturs sterkeste bilder, følelser, og forestillinger.

Maskinene blir mer som mennesker, menneskene blir mer som maskiner: sammenblandingen er grunnet i en dyp fascinasjon med maskiner og intelligens, og hva disse potensielt kan utrette.

Vi finner sammenblandingen ‘menneske-maskin’ igjen i nylige visjoner om kunstig intelligens, hvor forestillingen forsterkes, fordypes og intensiveres. Sjefen for Google DeepMind (Google’s KI-forskningsavdeling) forteller hvordan vi har ‘fått sand til å

tenke'. Sand blir til silikon, som blir til computer-chips, som prosesserer store mengder tall og data, som inngår i statistisk (maskin-) læring. Maskin-læring gjenkjenner mønstre i språk, og benyttes i store språkmodeller, som Gemini, ChatGPT, Claude, Mistral og DeepSeek. Men påstanden om at de 'tenker' i egentlig forstand er utvilsomt en projisering, en projisering som i dag blir strategisk anvendt av noen av vår verdens mektigste aktører

Før vi gir oss i kast med dagens diskusjoner omkring kunstig intelligens skal jeg gjøre en liten omvei, inn i hermeneutiske og samfunnsvitenskapelige analyser av vitenskap og teknologi. Disse vokste frem samtidig med en mer systematisk og strategisk fokus på visjon og fremtid som politiske styringsverktøy.

Filosofen Charles Taylor (noen ganger sammenlignet med Skjervheim) innførte et begrep om mennesket som et selvfortolkende dyr, og et tilsvarende begrep om forestillinger (the social imaginary). Med dette sikter han til bilder og begreper som deles av mange og er sentrale for kollektiv selvforståelse. Begrepet har blitt videreutviklet av vitenskapsteoretikere som Sheila Jasanoff og Brian Wynne for å forstå hvordan vitenskap og teknologi i økende grad inngår i vår kollektive forestillinger og talemåter. Begrepet er en versjon av sosiologiske analyser av løfters og forestillingers, og hype eller opphaussing, og deres sentrale strategiske rolle i vitenskap, teknologi, innovasjon og markeder. (løftenes, forventningenes og opphaussingens sosiologi)

Slike forestillinger eller visjoner er fokusert på tekniske objekter som gener, atomer eller språkmodeller. Men de er samtidig også visjoner for en fremtidig samfunnsmessig orden. De tilsynelatende tekniske visjonene er dypt investerte med ønsker, verdier og forestillinger om samfunnet, og hvordan det bør utvikle seg (feks atomkraft). Midt i det tilsynelatende tekniske universet finner vi derfor svært menneskelige meninger, ønsker og ambisjoner: om vår plass i verden; hvem vi er, og (ikke minst) hvem vi kan og bør bli i mulig fremtid.

BILDE 4. For å gi et eksempel på visjonenes instrumentelle og strategiske rolle: rundt årtusenskiftet introduserte EU-kommisjonen et forskningsprogram, hentet fra Silicon Valley, om 'intelligente omgivelser' ('ambient intelligence'). Ekspertgruppen som utformet visjonen (*Information Societies Advisory Group*), var eksplisitte på at deres mandat var å 'skape og forme fremtiden'. De formulerte en visjon som skulle fungere som felles siktemål, for å samle aktører på tvers av forskning, innovasjon og markeder. Visjonen var som følger:

Mennesker er omgitt av intelligente, intuitive grensesnitt som er innebygd i alle slags objekter, og et miljø som er i stand til å gjenkjenne og reagere på tilstedeværelsen av forskjellige individer på en sømløs, diskret og ofte usynlig måte (ISTAG, 2001).

Visjonen ble ledende for utformingen av seinere digitale forskningsstrategier i Europa, som tingenes internett, smarte teknologier, robotikk og kunstig intelligens. Men mest interessant: den ble prøvd ut i praksis, i Phillip Research daværende *Opplevelses-laboratorium* i Eindhoven. Forskerne bygde kunstige miljøer: boliger med intelligente vegger og apparater, helseklinikker, hele handlegater og butikkvinduer. Og, de inviterte ‘vanlige folk’ til å prøve disse miljøene ut, og til å bo i dem. Problemet som møtte de visjonære entreprenørene (fra Phillips, Silicon Valley og EU) var at deltakerne i eksperimentene ikke likte de nye omgivelsene, noe de ga tydelig uttrykk for. De ville ikke leve i en verden dominert av maskin- og forbruker-sentrerte forestillinger. De avslørte den opprinnelige visjonen som reduksjonistisk og mangelfull (og også som noe av et personvern-mareritt). Dette ble i noen grad også innrømmet av forskerne og ingeniørene selv. Fra dette kan vi lære flere ting, som er tematisert gjennom løftenes, forventningenes og opphaussingens sosiologi:

Vitenskap og teknologi utgjør i dag de sterkeste kildene for utformingen av visjoner for kollektiv fremtid. Disse visjonene er ikke i egentlig forstand prediksjoner (som om de gjerne inneholder elementer av prediksjon). De er, som noen sosiologer og filosofer sier, performative. Deres primære funksjon er ikke å utsi sannhet, men å overbevise, begeistre og mobilisere. Alle som har sett en Ted-talk, eller en glødende inspirert tale om innovasjon, forstår hva dette innebærer. Innovasjons-tale er gjennomsyret av løfter om revolusjon og ‘disrupsjon’ av eksisterende institusjoner; ikke gjennom politisk mobilisering, men gjennom teknologi. Om å bryte med det bestående, men også om å skape nye ‘synergier’ i nettverk, overkomme siloer og barrierer, om å skape det nye: en ny verden, nye markeder, nytt samfunn (og kanskje nye mennesker). Den sentrale aktøren i innovasjons-talen, er entreprenøren: en visjonær og karismatisk person, gjerne med noen grad av teknisk og markeds-innsikt.

I det som har blitt kalt den digitale nettverks- og plattform-økonomien, er et sentralt motiv *oppskalering* gjennom å involvere stadig flere aktører, prosesser og ting i de digitale plattformene. Økonomer snakker om nettverkseffekter hvor eieren av en plattform klarer å innrullere stadig flere brukere og deres data. Løftene om digitale nettverk og plattformer og de fantastiske fremtidene de vil bringe oss, er instrumenter i dette arbeidet.

Denne typen dynamikk var virksom i tilfellet med ‘intelligente omgivelser’. Selv om feilene i noen grad ble anerkjent av sine opphavsmenn ble prosjektet ikke forlatt, men snarere omdefinert og oppskalert, av visjoner om store data, smarte teknologier,

tingenes internett, roboter og kunstig intelligens. Disse inngikk også i opprettelsen av det digitale indre marked, virksomt fra 2015. De digitale visjonene ekspanderte dermed: fra en forsknings- og innovasjons-setting, over mot hele markeder og stor-skala infrastruktur, og institusjonell implementering. Vi kan til og med si: de inngår i en overordnet forestilling om Europa som grunnlagt på digital infrastruktur og digitale markeder, nylige formulert i EU som *AI Continent*.

De digitale visjonene hørte en gang til et relativt vel definert område kalt ‘forskning og innovasjon’, begrenset til universiteter, forskningsinstitutt, deler av industrien og forskningsprogrammer. Visjonene har nå beveget seg ut over disse grensene, og blitt aktive premissleverandører for politikk og næringsutvikling. Det altoverskyggende tilfellet av dette er kunstig intelligens, og ingen sted er det mer fremtredende enn i Trump-administrasjonen, hvor tek-oligarker som Elon Musk og Peter Thiel innehar sentrale posisjoner. Men også i Norge ser vi hvordan regjeringen har sluttet seg til disse visjonene gjennom sin KI-strategi, ved å love at 100% av offentlig sektor skal benytte KI innen 2030, og gjennom visjonen om at ‘Norge skal bli verdens mest digitaliserte land’.

BILDE 5. Når ChatGPT ble lansert i 2022 spredte forestillinger seg om hva denne språkmodellen kunne utrette, som ild i tørt gress. Modellen ble først sendt direkte til brukerne, som ble presentert for det etter hvert velkjente bruker-grensesnittet: hvordan kan *jeg* hjelpe *deg*? (=antropomorfisering). Tidligere økonomisjef i Open AI, Denise Dresser, uttalte hvordan hun *aldri (har) sett dette nivået av overbevisning spre seg så raskt og konsekvent i bransjen*. Med ‘bransjen’ siktet hun til store selskaper, og kappløpet om å inkorporere KI i virksomheten før alle andre. Finansbransjen, og alle større selskaper, utviklet raskt strategier for å inkorporere KI i virksomhetene.

BILDE 6. Med denne spredningen har diskusjoner som tidligere var forbeholdt relativt begrensede grupper av eksperter, om KIs farer og muligheter, spredt seg i offentligheten. Scenariene er både utopiske og dystopiske: fra KI skal redde verden, til at KI utsletter verden. Felles for disse er imidlertid en forståelse av denne teknologien som eksepsjonelt kraftig.

Sentralt i KI-forestillingen står begrep om aksellerasjon, hurtig omstilling, og kappløp. På et seminar om digitalisering av energisektoren, hørte jeg vår minister for digitalisering utsi at: *for å vinne i det globale kappløpet, må vi vinne teknologikappløpet. For å vinne teknologikappløpet må vi vinne kappløpet om kunstig intelligens*. Dette budskapet fremsettes med stort trykk av leder for oljefondet, Nikolai Tangen: *KI er det viktigste som skjer i verden nå Vi står midt i en teknologisk revolusjon. Alt går raskere med KI. Det er en superhjelper. Alle i Oljefondet bruker KI nå, og vi sparer flere millioner på det. I Norge vil KI kunne få opp effektiviteten og lønnsomheten*. På dette viset

fremstår KI som ikke bare ønskelig, men som nødvendighet, fordi det utgjør (den eneste farbare) veien inn i fremtiden. Dersom vi vil løse problemer knyttet til klimaendringer, et aldrende samfunn, konkurransekraft, migrasjon eller grunnleggende forskningsspørsmål: alt fremstilles som om det avhenger av at vi implementerer KI, snarest og på tvers av sektorer. Fremtiden er her nå, og de som ikke stiger om bord på KI-ekspressen vil tape i kappløpet, falle bakpå eller utenfor.

Men la oss forfølge KIs symbolunivers noe videre. Jeg vender da kort tilbake til hvordan vi møter KI-modellene gjennom brukergrensesnittet.

BILDE 7. Modeller som Chat GPT presenterer seg som universelle hjelpere: *spør meg om hva som helst* (= antropomorfisering). De som har brukt disse modellene en del har kanskje lagt merke til at de er akkompagnert av et forbehold: *Vær oppmerksom på at KI kan gjøre feil*. Det er vel kjent at modellene gjør feil, gir ubalanserte fremstillinger, og finner opp ting (de 'genererer'). Når de anvendes i den fysiske verden forårsaker KI-modellene ulykker, for eksempel i form av selvkjørende biler. Nettstedet *The AI Incident Database* er opprettet for å dokumentere slike hendelser.

KI-modellene er maskiner, og kan som sådan ikke tilskrives juridisk eller moralsk ansvar (= slutt på antropomorfisering). Mens den lille setningen 'KI kan gjøre feil' kan fremstå som en innrømmelse av modellenes begrensninger, er den også et forbehold, tatt av KI-selskapene, om at de ikke vil gjøres ansvarlige for eventuelle feil. Det er det i all hovedsak utviklerne, brukerne, og organisasjoner lenger nede i 'verdikjeden', som er.

KI-selskapene har fremmet historien om at modellene deres utgjør nødvendig utvikling og innovasjon, og må implementeres snarest. Men de samme aktørene nekter også å ta ansvar for ulykker forårsaket av KI. Det er en merkelig asymmetri, der entreprenørene får skape nye risikoer og farer, mens vanlige folk og det offentlige må ta konsekvensene når ting går galt. Teknologi-kritiker Cory Doctorow sammenligner KI med en rask bil. Men denne spesielle bilen er laget uten bremses. Det kan hende at rettsavgjørelser og forsøk på å regulere teknologien, som EUs AI-forordning, gradvis kan sette bremsene på. I øyeblikket blåser vinden i andre retninger: under påtrykk fra tek-oligarker og hjemlig industri er EU i ferd med å vingeklippe sitt eget regelverk.

Asymmetrien, i hvem som kan skape farer og risiko, og hvem som må leve med dem, passer godt inn i de digitale entreprenørenes ideologiske univers. Denne ideologien ble først beskrevet av Barbrook og Cameron som *The Californian Ideology* (90-tallet). Et annet begrep satt på den (av filosofen Langdon Winner), er tekno-liberalisme. BILDE 8. Da Internet vokste frem som globalt fenomen på nitti-tallet, sammenfalt dette med optimistiske idealer om frigjøring fra tradisjonelle hierarkier. Tekno-liberalerne forestilte seg, og forespeilte oss, et samfunn organisert rundt horisontale nettverk hvor alle kan

kommunisere med alle, og hente inn informasjon fra over alt. Etterhvert som store selskaper inntok internett (utover 2000-tallet) begynte det liberalistiske universet å knake i sammenføyningene. Vi fikk begreper om overvåkningskapitalisme (hvor person-sensitive data utgjør den sentrale valuta), og en gryende platformkapitalisme, monopoler kontrollert av stor-skala selskaper som Google, Amazon, Facebook, Palantir og Apple.

Gradvis har ideen om teknologisk skapt frihet endret seg. Palantir's leder og grunnlegger av PayPal, Peter Thiel, uttalte allerede i 2009 at frihet og demokrati ikke er compatible, og at demokratiet derfor må vike. Med frihet mente han frihet for teknologien og for entreprenørene: til å innovere uten hindre fra regulering eller mostridige borgere. Thiel's verste mareritt er en verden der selskapet hans underlegges regulering og må finne seg en posisjon innenfor planetens grenser for bærekraft. Han omtaler derfor Greta Thunberg som Anti-Kristus (men det er Gretha som står sammen med palestinerne på bakken, mens bombene regner over Gaza fra oven, hjulpet av Thiel og Palantir's algoritmer).

Peter Thiel har relativt nylig plassert det hele i en kristen-konservativ språkdrakt: mennesket er, etter skapelsesberetningen, ufullkomment. Våre vestlige samfunn har blitt dekadente, de har stivnet og forfaller: vi (i Vesten) produserer ikke lenger grensesprengende teknologier og vitenskap, økonomien og samfunnet stagnerer. Det er de digitale entreprenørenes rolle å forbedre menneskene og samfunnet gjennom teknologisk disrupsjon og revolusjon, dersom vi skal unngå stagnasjon og totalitarisme. Det er entreprenørens frihet som fremmes gjennom visjonen om frihet for teknologien.

Man må gjerne innvende at Peter Thiel, som vi også hørte om i forrige foredrag, er et særtilfelle selv for tek-entreprenørene og Silicon Valley, og i noen grad er det sant. Men alle de største KI-selskapene, og tek-gigantene som sponser og bruker modellene deres, bekjenner seg til såkalt Kunstig Generell Intelligens (KGI) som overordnet mål. KGI beskriver visjonen om en overmenneskelig, teknologi-skapt intelligens. Bill Gates formulerte det som følger (i 2023):

Når utviklere kan generalisere en læringsalgoritme og kjøre den med datamaskinens hastighet [...], vil vi ha en utrolig kraftig KGI. Den vil kunne gjøre alt en menneskelig hjerne kan, men uten noen praktiske begrensninger på størrelsen på minnet eller hastigheten den opererer med

BILDE 9. Jeg ba Copilot om å hente inn visjoner fra Internett (den er god til det!) av KGI. Den misforstod meg, og genererte i stedet sin egen versjon (dog basert på tilgjengelig materiale). Dette er ikke en eksakt representasjon av enkelt-personers ideologi eller

oppfatning, men bildet gir (i konsentrert form) en slags oppsummering. Det fremstiller en nær guddommelig kraft. Denne kraften, som er rein intelligens, er skapt i menneskets bilde men har overkommet menneskelige begrensninger og blitt til en gud. Selv om denne guden ikke finnes, annet enn i en forespeilt fremtid, utgjør denne fremtiden mye av grunnlaget for kravene til samfunnsmessig endring her og nå. Entreprenørene henter autoritet, fra fremtiden, fra en mulig overmenneskelig kraft som de lover vil inntreffe fordi de selv skal skape den. De manifesterer en form for hva Max Weber kalte karismatisk autoritet (ikke juridisk-regulativ):

Karismatisk autoritet, representert av profeten, er den reneste formen for autoritet ved at den krever retten til å bryte gjennom alle normative strukturer ... Så lenge profeten beholder sitt karisma, kan han ødelegge gamle normer og skape nye (Martin Spencer, Weber on legitimate norms and authority)

Vi snakker altså ikke bare om teknologiske artefakter, ting, objekter, her; KI er i høy grad også investert i en ytterliggående ideologi, som grenser til religion (og har gjort det lenge, jfr Wiener's bok).

Stilt overfor kraften, kompleksiteten og autoriteten i de digitale visjonene er det lett å føle maktesløshet, som mennesket på bildet stilt overfor den allmektige teknoguden. Her er tiden for å vende tilbake til Skjervheim og det instrumentalistiske mistaket: han advarte mot objektivisering av en selv og av andre. En ny type aktør, digitale maskiner, har entret samfunnet og på mange måter forsterket de instrumentalistiske trekkene som Skjervheim identifiserte. Om vi kun ser denne aktøren som en deterministisk maskin som vi ikke har noen innflytelse over, blir det nærliggende å skru Skjervheims analyse av objektivisering og instrumentelle mistak opp, med ganske mange hakk. Like gjerne som å spørre 'hvor er det instrumentalistiske mistaket', kan vi spørre: hvor er det *ikke*? De digitale maskinene er tilsynelatende over alt, hele tiden. Men dette blir for pessimistisk.

Men vi kan jo faktisk nøytralisere disse visjonene og forestillingene (og vi kan fortsatt ha KI, om enn i annen form). Vi kan avvise tek-entreprenørens karisma, og la være å følge visjonene (som i The Matrix: rød pille eller blå pille?). Det er her kritiske sosiologiske studier av opphaussing, av løfter og visjoner, har en viktig misjon: de kan vise frem det historisk betingede, filosofisk og vitenskapelig begrensede i teknologien, og i løftene og forestillingene som omgir den. Men enda viktigere: større deler av befolkning og offentlighet må (og, jeg tror: vil) innse, at denne utviklingen ikke er i vår interesse; se på debatten om digitalisering av skolen; se på den gryende debatten om plassering av datasentre.

Om vi innser at utviklingen ikke skjer i henhold til nødvendige lover, men er kontingent og kan påvirkes, vil mye være oppnådd. Andreas snakket i går om objektivisering av en selv og

andre som flukt fra frihet. En annen form for flukt fra frihet er å tro blindt på KI-gudens universalitet og nødvendighet, og entreprenørens karisma. FOR:

BILDE 10. KI-Maskinen er ikke et objekt!

Bare søk i vei, på internett, i faglitteraturen eller lovgivnings-litteraturen: ingen kan si nøyaktig *hva KI er*. Vi har en hang til å forestille oss KI som en slags deterministisk maskin; eller et slags menneske. Men den er ingen av delene. Maskin-læring og algoritmene er i noen grad er fleksible, og kan i noen grad tilpasse seg sosial kontekst, språk og menneskelig fellesskap. Teknologien utvikler seg i samspill med mennesker. Den former oss, men vi former også den.

Relatert: Copilot-versjonen av KGI overså en ting, nemlig alle de andre menneskene (ikke bare entreprenøren!) som er involvert i de store språkmodellene: programmerere, folkene som trener modellene (ofte outsourcet til land med billig arbeidskraft), designere, de som legger ned fiberoptiske kabler, bygger datasentrene, og flere til. KI sitter på toppen av gigantisk digital, materiell og menneskelig infrastruktur, driftet av mennesker.

Også relatert: KI fungerer bare gjennom tilgang til enorme mengder menneske-produsert materiale, fra bilder på Facebook til verdenslitteraturen, til nyheter, blogger og vitenskapelige og tekniske journaler. Det vil si: KI konsumerer enorme mengder menneskeskapt materiale, bygd på arbeidet til de mange innholdsskaperne som har levert materiale som leverer data til modellen (det er mange av oss i denne salen!). Tek-oligarkene hevder at dette materialet tilhører *dem* (plutselig er ikke intellektuelle eiendomsrettigheter viktige lenger: de vil ha tilgang til *alt*). Men dette er en tilraning. Uten det menneskeskapte innholdet, som tilhører oss alle (vitenskap, kunst, litteratur, brukererfaringer, nyhetsblogger, poster på sosiale medier, etc.), er KI i sin nåværende versjon, nær verdiløs.

Vi kan, for å parafrasere Skjervheim, gå fra passive tilskuere i en utvikling svært få av oss egentlig ikke ønsker, til aktive deltakere som former utviklingen, og skaper andre fremtider enn dem forespeilt gjennom de store KI- og teknologi-selskapene. Oppgaven er dels å beskytte en felles verden, som vi finner det artikulert hos Skjervheim. Men i dag må vi gå lengre: muligens fordi det instrumentalistiske mistaket har spredt om seg og vi er i en verre situasjon enn da Skjervheim skrev sine tekster. Men det er også dette: teknologi handler ikke bare om diskurs, offentlig debatt og bruk av ulike former for abstrakt kunnskap; teknologien skaper verdener. Mot-strategier til den dominerende og fremmedgjørende tek-ideologien, må også evne å gjøre dette: sette frem konkrete strategier for hvilke type samfunn vi ønsker, hvilke rolle maskinene skal ha i dette samfunnet, hvem som skal gjøre det og gjennom hvilke virkemidler vi ønsker å jobbe. Så, spørsmålet handler ikke bare om

ideologi og kunnskap; det handler også om materiell infrastruktur, maskiner og hvordan disse kan inngå i et menneskelig felleskap og en felles verden.
Takk for oppmerksomheten.

Digitale visjonar: hinsides instrumentalistiske mistak?

kjetil.rommetveit@uib.no

Skjervheimseminaret 2026

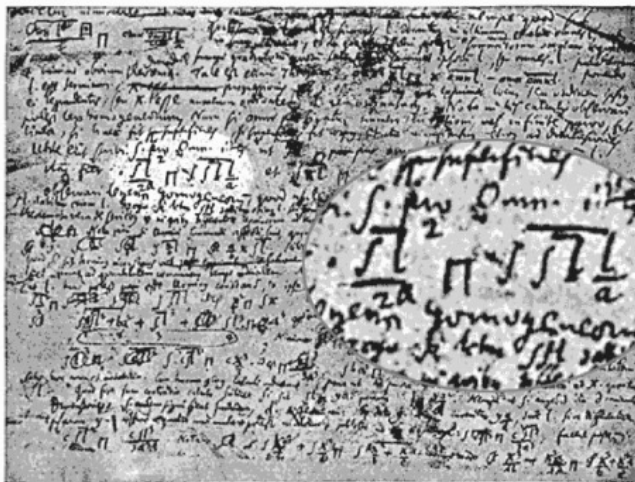
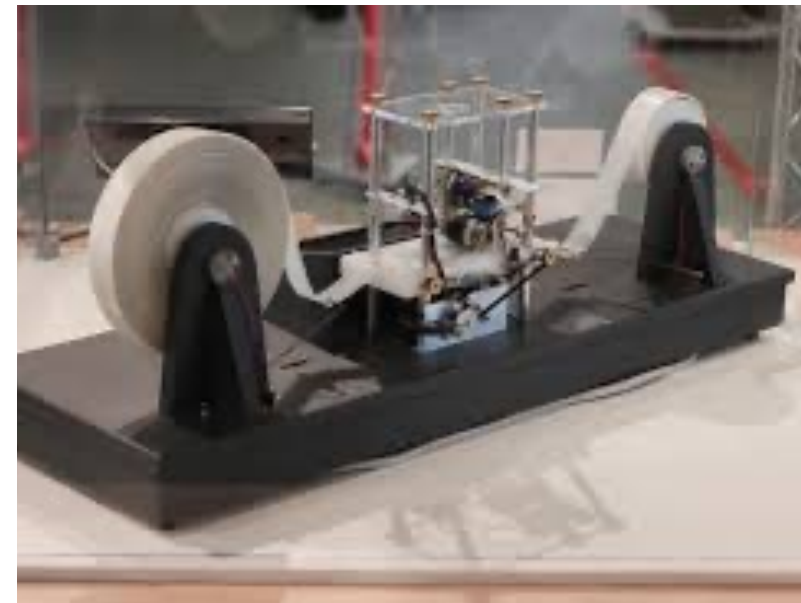
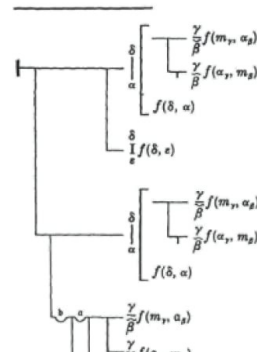
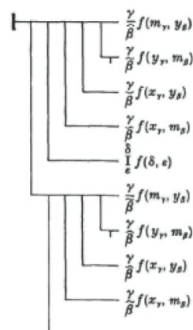


FIG. 124.—Facsimile of manuscript of Leibniz, dated Oct. 29, 1675, in which his sign of integration first appears. (Taken from C. I. Gerhardt's *Briefwechsel von G. W. Leibniz mit Mathematikern* [1899].)

Begriffsschrift

a formula language, modeled upon
that of arithmetic, for pure thought





**NORBERT
WIENER**

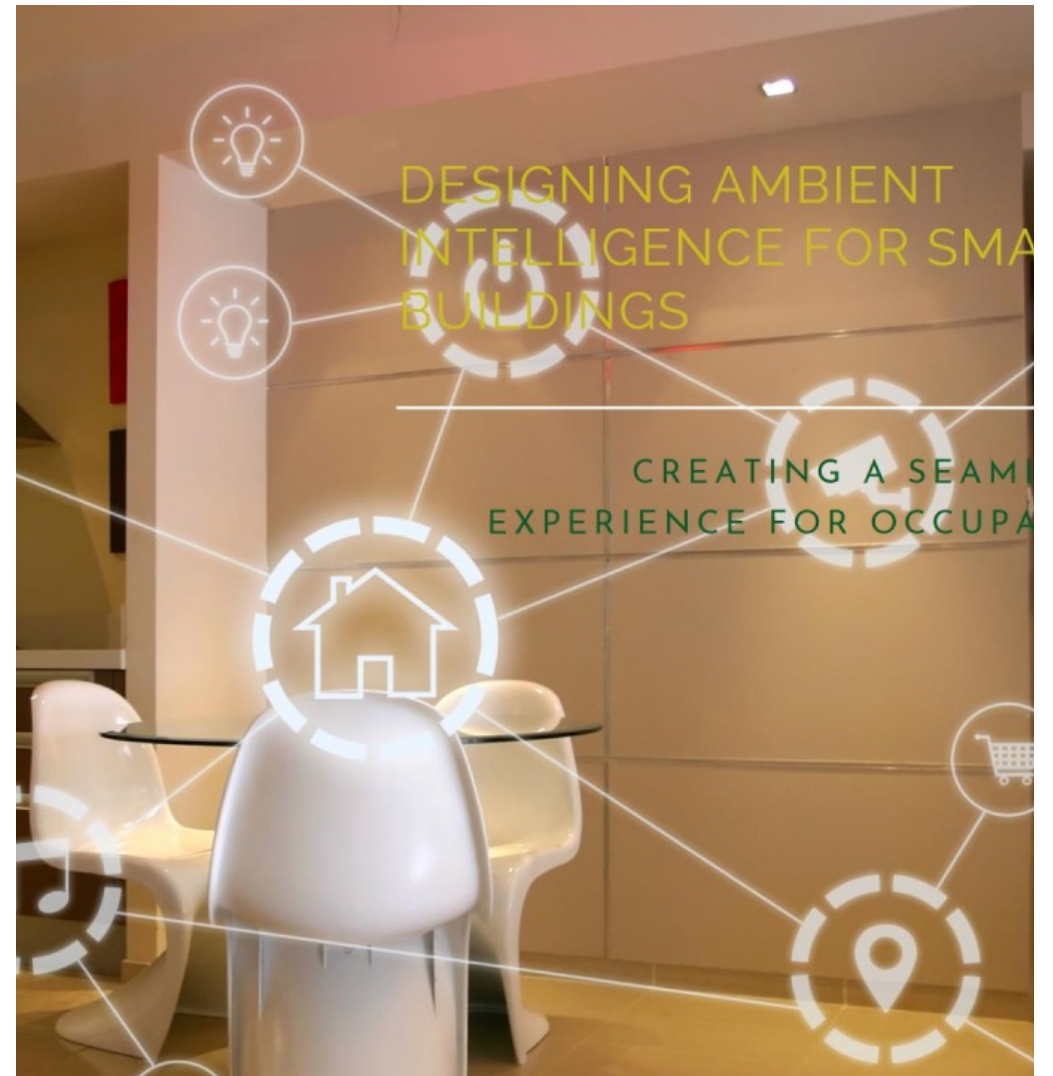
GOD & GOLEM, Inc.

A Comment on Certain Points where
Cybernetics Impinges on Religion



MIT 42
\$1.45





Velkommen, hvordan kan jeg hjelpe deg?

Send melding til Copilot



Learn

Find

Summarize

Suggest



Dario Amodei

Machines of Loving Grace¹

How AI Could Transform the World for the Better

October 2024

The list of positive applications of powerful AI is extremely long (and includes robotics, manufacturing, energy, and much more), but I'm going to focus on a small number of areas that seem to me to have the greatest potential to directly improve the quality of human life. The five categories I am most excited about are:

1. Biology and physical health
2. Neuroscience and mental health
3. Economic development and poverty
4. Peace and governance
5. Work and meaning



Send melding til Copilot



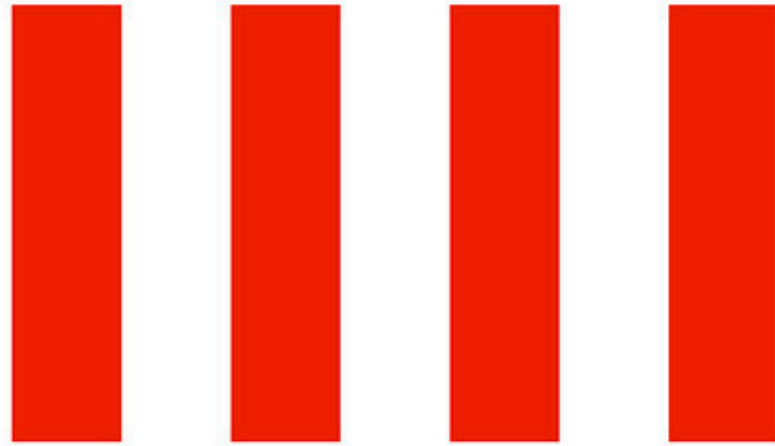
Copilot



Innhold generert av kunstig intelligens kan være feil

◆ AI-oversikt

Because artificial intelligence cannot be sued, legal and moral responsibility for AI errors ultimately falls on the humans and organizations involved. Accountability is typically distributed along the AI supply chain—from developers and data providers to the employers or professionals who deploy and oversee the technology. [← RiskInfo.ai +2](#)



Freedom to connect.
Freedom to learn.
Freedom from tyranny.
Freedom to innovate.



KI er ikke et objekt!