KI-fantastenes uberegnelige drømmer om beregnbar virkelighet

Fjasete høytenkning om kunstig intelligens' generelle metafysiske fordringer

Emanuel T. Frogner

Sammendrag

Fremtidsvyene om generell kunstig intelligens har satt verden i brann; KI-fantaster overbyr hverandre med himmel og helvete, Antikrist eller Guds komme. Er disse visjonene simpelthen en ekstrapolering av kapasiteten til dagens generative transformatormodeller, eller ligger det metafysiske fordringer bak drømmene?

Jeg er ingen ekspert på transformatorenes dunkle mulighteter til å bruke statistikk og grafteori for å generere Miyazaki-pastisjer til allment bruk. Jeg står heller ikke på noen av de steile frontene med endetidsprofeter som enten lover helvete eller himmel om den kunstige intelligensen blir generell nok. Jeg er og blir en slags forvirret dilettant i det hele. Men min amatørisme vis-à-vis modellenes velde er også en fordel: Den gir et skråblikk på alle lovnader, festtaler og skremselsbilder som kommer fra alle kanter. De kommer vel snarere stort sett fra to kanter: den politiske, i kraft av statsråders bedyrelse om at dette og hint skal effektivisere saksbehandlingen,[1] og den teknologiske, som nærmest ubegrensede visjoner om hva KI kan bety for fremtiden – enten med negative eller positive fortegn, forfektet av fantaster som Musk, Altman,

[1]Karianne Tung, «Naiv, sier du? 80 prosent KI i offentlig sektor er bare starten», Aftenposten, 14. mai 2025, https://www.aftenposten.no/meninger/debatt/i/bm5mlA/naiv-sier-du-80-prosent-ki-i-offentlig-sektor-er-bare-starten.

Bostrom eller Thiel.[2] Som kontrast kan jeg kanskje driste meg til å kalle min tilnærming filosofisk – selv om jeg som fagmenneske er ganske usikker på hva det egentlig betyr, men slike tvetydigheter kommer vel med terrenget.

Genererte modeller til tvetydig glede

Selv om noen fremstillinger av disse statistiske maskineriene synes litt merkelige for meg, er jeg ikke i tvil om at de har bruksområder. Som dyslektiker har de gjort meg mer i stand til å formulere tanker uten graverende ortografiske feil. I denne sammenhengen er det også åpenbart hvordan deres mekanisme løser problemet: de sammenligner simpelthen mitt dyslektiske uttrykk statistisk med det de har øvd seg på tidligere, og presterer dermed, med en viss grad av presisjon, å redde inn mine obskure svermerier og fjasete resonnementer – i hvert fall på det leksikale nivået.

Imidlertid er det noe annet som slår meg når jeg bruker de statistiske nevrale nettverkene til å rette mine kråketær: det virker som KI-en «avskyr» tvetydighet, særlig av den semantiske arten. Det kommer tydelig frem i at den prøver å «rette» tvetydigheten bort – redusere setningens flertydighet og erstatte dem med relativt bastante entydige uttrykk. Jeg er fanget i en filosofisk tradisjon, med det flamboyante tilnavnet dekonstruksjon, som nettopp har en tendens til å påpeke tvetydighet og motsetninger i både språk, verden og tenkning – og bruke det både kritisk og generativt (det siste blir ofte glemt av våre meningsmotstandere). Det at disse generative modellene ikke «liker» semantisk tvetydighet – når noe verken er på den ene eller den andre måten

[2]Catherine Clifford, «Elon Musk at SXSW: A.I. Is More Dangerous Than Nuclear Weapons», CNBC, 13. mars 2018, https://www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html; Samantha Kelly, «Sam Altman warns AI could kill us all. But he still wants the world to use it», CNN, 31. oktober 2023, https://edition.cnn.com/2023/10/31/tech/sam-altman-ai-risk-taker/; Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014); Tina Nguyen, «Silicon Valley's latest argument against regulating AI: that would literally be the Antichrist», The Verge, 25. september 2025, https://www.theverge.com/ai-artificial-intelligence/785407/peter-thiel-antichrist-tech-regulation; Will Douglas Heaven, «How AGI became the most consequential conspiracy theory of our time», *MIT Technology Review*, 30. oktober 2025, https://www.technologyreview.com/2025/10/30/1127057/agi-conspiracy-theory-artifcial-general-intelligence/.

gitt en klart definert ramme – er uansett i grunnen ikke så underlig. Hele KI-ens system er bygget på en formalisering av en enten–eller-logikk som kan fange alt som er beregnbart. Spørsmålet for dem som lover genererte gull og grønne skoger, eller en slags gammeltestamentlig avgud av det kunstige slaget, er hvorvidt menneskelig bevissthet og intelligens er beregnbar på samme måte.

Beregnbarhet som målestokk

Alle nåværende datamaskiner, samt med høy sannsynlighet alle de fremtidige, veldig kunstferdige og intelligente systemene, er i prinsippet såkalte Turing-maskiner. Formaliseringen, ført i pennen av matematikeren og logikeren Alan Turing, er stadig det alle datamaskiner opererer etter. Det Turing viste, var at alle beregnbare problemer kan, med nok tid, løses av en enkel liten dings som leser input og utfører operasjoner steg for steg, til den enten gir et resultat eller stanser. Men Turing viste også at det finnes en grense for hva som er beregnbart. Dette er kjent som halting-problemet.[3]

Halting-problemet kan forklares slik: Gitt en vilkårlig datamaskin og et vilkårlig program med et vilkårlig input – kan vi på forhånd avgjøre om programmet kommer til å stoppe (haltere), eller om det vil fortsette å kjøre i det uendelige? Turing beviste at det ikke finnes noe generelt program som kan avgjøre dette for alle mulige programmer og inputs. Det betyr at selve spørsmålet om et program stanser eller ikke, i mange tilfeller ikke lar seg beregne.

I forlengelsen av dette dukker et annet spørsmål opp: Er verden beregnet i seg selv eller kan den beskrives av matematikken? Eller for å si det på en annen måte: oppdager vi reglene for naturen i våre matematiske modeller, eller beskriver vi fenomener med en høy grad av presisjon og forutsigbarhet ved bruk av matematikk – noe som innebærer at vi allerede før vi bruker

[3] A. M. Turing, «On Computable Numbers, with an Application to the Entscheidungsproblem», *Proceedings of the London Mathematical Society*, 1937.

matematikkens modeller må ramme inn det vi undersøker på et slikt vis at det er meningsfullt å snakke om det i kvantitative størrelser i utgangspunktet?

Dette spørsmålet kan enkelt illustreres med geometri: Ingen sirkler vi møter i den opplevde (fenomenelle) virkeligheten lever helt opp til den matematiske konstruksjonen av en sirkel. Den opplevde sirkelen er rufsete i kantene, for å si det sånn. Spørsmålet er om vi bruker de matematiske konstruksjonene til å forstå de erfarte sirkelene. Følgelig kan vi bruke dem og forstå dem på forskjellige måter som er veldig nyttige – som avdekker strukturelle sammenhenger, eller fungerer som en slags abstrakt analog til det som skjer. Eller finnes det snarere, skjult bak eller i den opplevde sirkelen, matematiske forhold som er mer virkelige enn den opplevde sirkelen – forhold som den eksisterer på bakgrunn av? Den siste holdningen kalles gjerne platonisme i matematikkfilosofien, fordi den minner om formlæren assosiert med Platon.

Livets uberegnbarhet

I mot en slik platonistisk holdning vil jeg innvende: Selv gitt et algoritmisk syn på virkeligheten, opererer levende systemer alltid innenfor fluiddynamiske sannsynligheter og systemteoretisk selvforming. De er sågar i sin konstitusjon fulle av tilbakekoblingsmekanismer og omgitt av væsker og gjennomstrømmes av dem.[4] Mange slike prosesser er formelt kaotiske: Vi kan kjenne den opprinnelige tilstanden, men likevel ikke beregne utfallet. På opplevelsesnivået er det også en kjensgjerning at levende systemer som sanser, må håndtere og forvalte motstridende inntrykk i en sammensatt omverden – der mediering av tvetydighet antagelig har større betydning enn beregnet sannsynlighet.

I tilbakekoblingen mellom opplevelsen og resten av de levende systemene finner vi ikke bare en analogi, men en formell likhet med den uendelige løkken som oppstår i halting-problemet: Levende systemer lar seg ikke avgjøre. Deres plastiske beskaffenhet gjør at det eneste levende systemet som kan bestemmes avgjort, er et dødt levende system. Følgen av dette er at betingelsen for

[4]Raymond Noble og Denis Noble, *Understanding Living Systems* (Cambridge: Cambridge University Press, 2023).

levende systemer er tvetydighet og uberegnbarhet – noe som i grunnen er motsatsen til det KI er.

Igjen vil jeg understreke at det er mye algoritmer og formelle logiske systemer implementert i elektriske kretser kan løse – både mer effektivt og bedre enn våte sekker av membraner og tilbakekoblingsmekanismer – men håndteringen av tvetydighet er ikke en av dem. Gitt beregningen av tvetydighet vil KI-en bare fastsette en statistisk distribusjon og få et entydig uttrykk ut av den i form av en probabilistisk beregning. Imidlertid er tvetydigheter noe som må håndteres og medieres, ikke beregnes. Videre er mange av de oppgavene som vi ønsker å bruke KI til – fra saksbehandling og å lette ensomheten[5] til å løse verdens gjennomgripende avveininger – oppgaver som bør betegnes som uberegnelige håndteringer av tvetydige og bråkete inntrykk. KI-en, på sin side, reduserer støyen med en statistisk graf og oppnår følgelig noe mer entydig, men uten noen vurdering av entydighetens sannhetsgehalt. Følgelig, om problemet handler om å håndtere tvetydighet, har vi bare sett bort fra problemet snarere enn å ha løst det om vi overlater det til svulmende transformatormodeller.

Generative språkmodeller og språkets generativitet

Hvorvidt virkeligheten og livet i bunn er beregnbare, er i siste instans et empirisk spørsmål – bak livets uberegnbare tvetydighet kan det muligens ligge en entydig og beregnbar virkelighet. Likevel synes det som at mange av dem som predikerer en algoritmisk gud – enten som dommer eller som frelser – allerede har konkludert: virkeligheten er i dypeste forstand et beregnbart fenomen. De som hevder at vi lever i en simulering, gjerne satt ut i live av vår høyteknologiske etterkommere,[6] tar dette synet til sitt ytterpunkt; virkeligheten er allerede et stort regnestykke utført av en Turing-maskin.

[5]Meghan Bobrowsky, «Zuckerberg's Grand Vision: Most of Your Friends Will Be AI», The Wall Street Journal, 7. mai 2025, https://www.wsj.com/tech/ai/mark-zuckerberg-ai-digital-future-0bb04de7.

[6]Nick Bostrom, «Are You Living In a Computer Simulation?», *Philosophical Quarterly* 53, nr. 211 (2003).

Imidlertid er det kjent at det finnes formelt definerte grenser for beregnbarhet. Det sentrale spørsmålet blir derfor hvilke fenomener som faller innenfor det beregnbare, og hvilke som ikke gjør det. Enn så lenge mener jeg vi har liten grunn til å tro at menneskelig bevissthet er et slikt beregnbart system. Jeg er endog en av dem som mener at selve Væren er tuftet på tvetydighet; muligheten for realiseringen av en generell intelligens som overgår oss på områdene vi er gode på – å håndtere tvetydighet – innenfor et deterministisk logisk system, virker heller liten.

Dette er bøygen for en generell kunstig intelligens: Den må på én og samme tid overgå levende systemer i å håndtere tvetydighet og bevare sin stringente logiske beregnbarhet. Om det er mulig å gjøre innenfor et beregnbart system, er avhengig av at den opplevde tvetydigheten egentlig er et uttrykk for en underliggende entydighet og videre at alle relevante problemer til syvende og sist er beregnbare.

Det er mye som står åpent vis-à-vis hva som er beregnbart eller ei. Imidlertid, som teknologi for å løse problemer den er egnet til, det være seg proteinfolding og analyse av røntgenbilder, for ikke å glemme ortografisk retting av dyslektiske kråketær, er statistiske transformatormodeller et utsøkt verktøy. Når det er sagt er fellestrekke ved disse eksemplene at en probabilistisk tilnærming står til problemet.

Videre vil jeg understreke at effektene som kommer til syne i store språkmodeller, antagelig har mer med språkets generative struktur å gjøre enn med modellenes kapasiteter. I KI-ens vidunderlige kraft ligger det noe langt eldre enn de formaliserte Turing-maskinene, kanskje like gammelt som vår språklige kapasitet: at språket kan virke på seg selv gjennom bruk og uttrykk. Samtidig er det mulig at noe nytt trer frem idet språket maskinelt podes på seg selv – at språkets «generelle plastisitet», dets evne til å formgi seg selv, blir synlig på nye måter i statistiske generative språkmodeller. Som Cathrine Malabou nylig har påpekt, er denne avenyen ennå lite utforsket.[7] Malabous tematisering av store språkmodellers betydning for bevissthet og språklig

[7]Catherine Malabou, «Dear ChatGPT...» (EGS Podcast, 30. oktober 2025), https://pact.egs.edu/podcasts/dear-chatgpt/.

eksistens tilhører et mye mer nærliggende virkelighetsnivå enn drømmen om en generell kunstig intelligens.

Uberegnbare drømmer om beregnbar virkelighet

Inntil videre må spørsmål rundt språkets egen beskaffenhet bli liggende, og min foreløpige konklusjon blir: Det de platonske KI-fantastene, med sine drømmer om algoritmiske avguder og forfedresimulasjoner, kommer til torgs med, er en politisk visjon. I den skal livet bli bestemt av systemer som antagelig formelt ikke egner seg til å fange livets tvetydighet.

I samfunnet tar denne oppfatningen form av en politisk handling som minner om de andre algoritmiske herjingene på nettet – først og fremst i sosiale medier – som i hvert fall må ha noe av skylden for at vi er blitt så dårlige til å håndtere tvetydigheter menneskene imellom. Videre har den materielle implementeringen av disse algoritmene miljømessige konsekvenser som er ganske entydige.[8]

Dermed illustrerer KI-fantastenes drømmer om en generell kunstig intelligens hvordan filosofiske problemstillinger stadig innhenter oss i hverdagslivet og går fra å være teoretiske spørsmål til en praktisk virkelighet som både påvirker hvordan vi håndterer ressurser i dag og hvordan vårt sosiale liv blir i morgen. Om KI-fantastene får heier fritt kan vi risikere at morgendagens samfunn blir basert på en idealistisk fordring, som har mer til felles med fabelprosaens spekulasjoner enn nøkterne undersøkelser av hva som er beregnbart eller ei. Uansett er gambiten forspeilet fra KI-fantastenes kant så dramatisk at, uavhengig av dens troverdighet, tvinger den frem maktpolitiske forordninger som fort kan bli et mareritt. For min del undres jeg over om den metafysiske spekulasjonen og den politiske visjonen egentlig er uttrykk for det samme:

[8]Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (New Haven: Yale University Press, 2021).

En verden som fullt ut kan gjøres rede for i modeller og systemer fantastene selv kontrollerer.

Slik fortsetter KI-fantastenes politiske handlinger å stake ut veien mot en generell, entydig politisk dumskap snarere enn å styrke intelligensen som kan peke ut retning i samtidens uoversiktlige tvetydighet. Til syvende og sist avtegner KI-fantastenes drømmer vår oppgave som et negativt bilde. Det er i det minste et slags tvetydig bidrag.

Bibliografi

Bobrowsky, Meghan. «Zuckerberg's Grand Vision: Most of Your Friends Will Be AI». The Wall Street Journal, 7. mai 2025. https://www.wsj.com/tech/ai/mark-zuckerberg-ai-digital-future-0bb04de7.

Bostrom, Nick. «Are You Living In a Computer Simulation?» *Philosophical Quarterly* 53, nr. 211 (2003).

———. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014.

Clifford, Catherine. «Elon Musk at SXSW: A.I. Is More Dangerous Than Nuclear Weapons». CNBC, 13. mars 2018. https://www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html.

Crawford, Kate. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press, 2021.

Heaven, Will Douglas. «How AGI became the most consequential conspiracy theory of our time». *MIT Technology Review*, 30. oktober 2025. https://www.technologyreview.com/2025/10/30/1127057/agi-conspiracy-theory-artifcial-general-intelligence/.

Kelly, Samantha. «Sam Altman warns AI could kill us all. But he still wants the world to use it». CNN, 31. oktober 2023. https://edition.cnn.com/2023/10/31/tech/sam-altman-ai-risk-taker/.

Malabou, Catherine. «Dear ChatGPT...» EGS Podcast, 30. oktober 2025. https://pact.egs.edu/podcasts/dear-chatgpt/.

Nguyen, Tina. «Silicon Valley's latest argument against regulating AI: that would literally be the Antichrist». The Verge, 25. september 2025. https://www.theverge.com/ai-artificial-intelligence/785407/peter-thiel-antichrist-tech-regulation.

Noble, Raymond, og Denis Noble. *Understanding Living Systems*. Cambridge: Cambridge University Press, 2023.

Tung, Karianne. «Naiv, sier du? 80 prosent KI i offentlig sektor er bare starten». Aftenposten, 14. mai 2025. https://www.aftenposten.no/meninger/

debatt/i/bm5mlA/naiv-sier-du-80-prosent-ki-i-offentlig-sektor-er-bare-starten.

Turing, A. M. «On Computable Numbers, with an Application to the Entscheidungsproblem». *Proceedings of the London Mathematical Society*, 1937.