

Fra AI til AGI i 2025?

Landvindinger indenfor AI i de seneste år kalder på eftertanke om hvordan AI vil udvikle sig i de kommende år.

Indenfor AI er det i høj grad de store sprogmodeller, som ChatGPT, der i de seneste år har trukket de store overskrifter. Modellerne forbedres hele tiden. Og 2024 var fyldt med overskrifter om at AI nu nærmer sig et menneskelignende niveau. Så starten af 2025 virker som en god anledning til at reflektere lidt over (del 1) de seneste AI landvindinger og prøve at komme det lidt nærmere (del 2) hvad der egentlig mangler for at AI bliver til AGI, menneskelignende intelligens.

Der er ikke længe til at alle telefoner bliver udstyret med et multimodalt AI program, der på baggrund af input fra telefonens kamera og mikrofon vil svare på spørgsmål og kommenterer verdenssituationen, det den ser, på engelsk, usbekisk og dansk. Næste skridt kunne meget vel være at få sådanne modeller kørende inde i humanoide robotter, der går rundt på 2 ben blandt os. I Nancy, Frankrig, kunne man i dette efterår på konferencen "Humanoids 2024" se mange af de robot modeller, der i de kommende år, ihvertfald ifølge producenterne, vil komme ud og hjælpe til på lagre, vise vej i storcentre, assistere indenfor sundhedssektoren, og skabe tryghed i bycentre ved at patruljere i gaderne om natten. Først nogle få robotter, så tusinder, og senere millioner af robotter¹.

Fra de selvkørende biler ved vi dog at der kan være langt fra en vision til et AI system, der rent faktisk kan løse en given opgave på niveau med et menneske. General Motors trak sig i december ud af robot taxa firmaet Cruise. Efter, siden 2016, at have investeret mere end 10 milliarder dollars i Cruise, og med udsigt til at skulle investere endnu mere for at få en forretning ud af det, kunne man på dem forstå, at springet fra hjælpe-systemer til en menneskelig chauffør, og så til at få fuldautomatisk taxaer på alle verdens veje måske alligevel var ret stort². Der er forskel på at køre ligeud på en motorvej i Californien og til at begå sig trafikken i New Delhi.

Det sagt, så har vi jo i de seneste år næsten dagligt fået nye beretninger om nye AI systemer, der er godt på vej til at opnå Artificial General Intelligence. Generel intelligens der kan forstå eller lære enhver intellektuel opgave, som et menneske kan løse. I december slog OpenAI³ igen til, og meddelte at deres nye "reasoning" modeller, har opnået en score på 87.5 % på en type af mønster genkendelses opgaver⁴, hvilket overgår en typisk menneskelig score, der ligger på 85%. Her taler vi ikke længere "bare" om store sprogmodeller, der trænet på enorme tekstmængder, kan give det mest sandsynlige svar (tekst) på et givet spørgsmål, i en given kontekst. I disse seneste "reasoning" modeller, benyttes forskellige teknikker til at undersøge modellens interne dialog og planlægge det bedste svar inden dette svar sendes videre til brugeren⁵.

Står vi lige foran det næste store AI gennembrud?

I et efterhånden velkendt mønster gik der ikke mange dage fra OpenAI præsentationen til at den amerikanske psykolog og AI forsker Gary Marcus endnu engang følte behov for at fortælle verden at sådanne generative AI systemer er gode til at give svar der lyder godt, eller plausibelt, i en given kontekst, men ikke til at forstå, på et dybere niveau, hvad de egentlig siger. De kan ikke fakta-tjekke sig selv. Hvilket så giver anledning til endeløse såkaldte "hallucinationer", hvor systemet påstår noget der ikke er rigtigt. Uden selv helt at forstå om et givet svar er rigtigt eller forkert. Ifølge Marcus kan man lave gode demoer med den slags generative AI systemer, og måske bruge dem til at brain-storme med, men disse systemer er ikke særligt gode komponenter i driftssikre produkter⁶. Marcus forsømte heller ikke at fortælle os at OpenAI står til at få et driftstab tab i 2024 på 5 milliarder dollars, hvilket i hans øjne gør at prisansættelse på firmaet, 80 milliarder dollars, er alt for høj⁷.

Det er kort sagt ret forståeligt at forbrugere og investorer ved indgangen til 2025 kan føle sig lidt forvirrede over hvad de egentlig skal tro. Står vi overfor et gennembrud med AI der bliver til AGI,

artificial general intelligence. Eller er vi udsat for "corporate hype", der overdriver hvad der er på hylderne, i håb om fortsatte investeringer i produkter, der ikke helt holder hvad de lover, og slet ikke genererer profit der berettiger investeringerne?

IT branchen har selvfølgelig været her før. Der har tidligere været perioder med store fremgange efterfulgt af nedgange indenfor AI, de såkaldte AI vintre. I 1973 udkom den såkaldte Lighthill rapport i England, bestilt af det engelske parlament, hvori James Lighthill slog fast at AI "fuldstændigt havde fejlet i at opnå sine grandiose mål", og "kun kunne bruges til at løse stærkt simplificerede legetøjsproblemer"⁸. En efterfølgende nedgang i investeringer i UK og USA gav færre AI projekter og derved en såkaldt AI vinter. I 1980'erne var der stor interesse for de såkaldte "vidensbaserede" systemer, som dog også viste sig at have svært ved at leve op til opskruede forventninger, hvilket så igen udløste en AI vinter med manglende vilje til at investere i AI systemer. For de der skulle have glemt det, kan det nævnes at selve ordet "AI" i 1990'erne, og i begyndelsen af det nye årtusinde, havde et så kontroversielt omdømme, at mange af os der blev uddannet indenfor IT og AI i den periode, af venligtsindede mentorer fik forklaret, at man i ansøgninger nok hellere skulle skrive at man havde arbejdet med ting som f.eks. Bayesian networks, kontrol systemer eller lignende ting, der lød mere "teknisk". Luftige ord som AI skulle man helst ikke bruge alt for mange af i ens ansøgninger.

Tingene ændrede sig for alvor da AlexNet, et neural net, designet af Alex Krizhevsky i samarbejde med Ilya Sutskever og (2012) nobelpristageren Geoffrey Hinton, i 2012 viste sig at være den bedste løsning til at kategorisere billeder i den såkaldte ImageNet Challenge. Og derfra er resten som bekendt historie, med det ene store AI gennembrud efter det andet, inklusivt frigivelsen af ChatGPT i November 2022⁹. Som redaktøren af Wired Magazine, Kevin Kelly, skrev allerede i 2014: "Det er let at forudsige hvad businessplanen er for de næste 10.000 start-ups: Take X and add AI"¹⁰.

Der arbejdes ihærdigt på at gøre de store sprogmodeller endnu bedre.

Det sagt er det selvfølgelig stadig interessant om de store sprogmodeller, som ChatGPT, der har løbet med megen af medie omtalen af AI (i tiden siden November 2022), om de virkelig er lige præcis den teknik der ikke bare giver en god løsning inden for et område, men om de er et skridt på vejen mod generel kunstig intelligens, AGI, sammenlignelig med menneskelig intelligens.

Før 2022 kunne man let støde ind i sprogmodeller der gav så mærkelige svar, at man ikke på nogen måde mistænkte dem for at være specielt intelligente. Modellerne var små og ikke trænet på specielt store mængder data. Og der havde ikke, efter træningen, været mennesker inde over modellerne og tilsætte "blokering" af uheldige, forkerte svar. Idag kan man genskabe sig oplevelsen ved f.eks. at køre modellen "llama2-uncensored" på ens egen laptop¹¹. Under min test her i denne vinter lykkedes det mig at komme ind i en samtale, hvor modellen påstod at den havde børn, Jack og Alex, som begge var mennesker. Tingene begyndte at blive endnu mere mærkelige. da jeg gerne ville vide mere om hvordan sprogmodeller kan få menneskelige børn, og fik at vide at det var en transformation, magen til den man kender når en drage bliver til et menneske.

Årsagen er selvfølgelig at man her taler med den "rå" sprogmodel. Der anvender "next-word prediction", hvor man får det tilbage der oftest følger i de tekster den er trænet på, med en vis sandsynlighed. Modellen er altså sikkert blevet trænet på en tekst fra internettet om drager. Uden helt at forstå, at vi faktisk ikke kender så mange eksempler på virkelige drager.

ChatGPT er jo selvfølgelig langt mere avanceret. Og forbedres hele tiden. Alligevel kunne man nu relativt nemt i foråret 2024 finde eksempler på spørgsmål hvor man ikke fik synderligt gode svar.

F.eks. kunne man i foråret spørge ChatGPT: "Forestil dig at du står foran 3 kasser med frugt. Alle kasserne har en etiket på dem med, hvilken slags frugt der er i kassen. en kasse indeholder æbler, en anden pærer, og den sidste kasse indeholder blandede frugter. Alle etiketter er forkerte med hensyn til, hvad der faktisk er i kassen. Uden at kigge, hvordan kan du tage en frugt fra en kasse og derefter sætte alle etiketterne korrekt?". Det rigtige svar er, at man skal tage en frugt fra kassen med blandede frugter. Hvis det var et æble, skal der på den kasse stå æbler. Etiketten hvor der stod æbler

må også have været forkert. Erstatte man den med etiketten blandede frugter bliver etiketten pærer ved med at stå forkert, så den etiket skal derfor nok erstattes med etiketten pærer. Osv.

Da ChatGPT i foråret rutinemæssigt svarede forkert på dette spørgsmål, var det selvfølgelig en stor opmuntring for alle der ikke helt var klar til AGI. Efter sommerferien var historien dog en anden. ChatGPT var opdateret og kunne nu pludselig svare på dette, og mange andre spørgsmål, den ikke tidligere havde kunnet svare på. Forbløffelsen var stor. Helt som da den undrende offentlighed i 2023 fik at vide at ChatGPT havde opnået en bestået score på visse amerikanske jura kurser¹².

En nærliggende mulighed var dog at ChatGPTs skabere, OpenAI, måske var blevet trætte af historier på sociale medier om ChatGPTs manglende evne til at svare på spørgsmål om frugtkasser. Og nu havde fået inkluderet de rigtige svar i teksterne der trænes på. Ligesom det også virkede sandsynligt at amerikanske jurastuderende havde diskuteret eksamensspørgsmål på sociale medier, som der herefter var blevet trænet på. Som mennesker også ved, når man selv skal til eksamen, er det lettere at huske svar, på en mindre mængde kendte spørgsmål, og gentage dem, end det er at forstå et helt felt, og på den baggrund svare på spørgsmål man aldrig har hørt før.

Jo, for AI verdenen var 2024 i sandhed et år med meldinger i øst og vest. I nogle rapporter stod vi lige foran store, afgørende gennembrud, mens andre mente at AI historien, fra 1950erne og frem, med bakker og daler, bare var ved at gentage sig, nu på vej mod en mere nøgtern realisme i 2025. Men i begge scenarier med krav til alle om at reflektere over situationen og arbejde omhyggeligt med at finde den bedste vej frem, udfra hvad man selv ved om ens egen verden.

- Dette var del 1 af "Fra AI til AGI i 2025".

Simon Laub

¹<https://www.forbes.com/sites/robtoews/2024/03/10/10-ai-predictions-for-the-year-2030/>

²<https://www.reuters.com/business/autos-transportation/general-motors-drop-development-cruise-robotaxi-2024-12-10/>

³<https://openai.com/12-days/>

⁴<https://arcprize.org/arc>

⁵<https://arstechnica.com/information-technology/2024/12/openai-announces-o3-and-o3-mini-its-next-simulated-reasoning-models/>

⁶<https://www.wired.com/story/generative-ai-will-need-to-prove-its-usefulness/>

⁷<https://finance.yahoo.com/news/report-claims-openai-burned-8-105046378.html>

⁸<https://www.brightworkresearch.com/first-ai-winter-and-what-the-lighthill-report-said-about-progress/>

⁹<https://openai.com/index/chatgpt/>

¹⁰<https://www.wired.com/2014/10/future-of-artificial-intelligence/>

¹¹<https://ollama.com/library/llama2-uncensored>

¹²<https://edition.cnn.com/2023/01/26/tech/chatgpt-passes-exams/index.html>