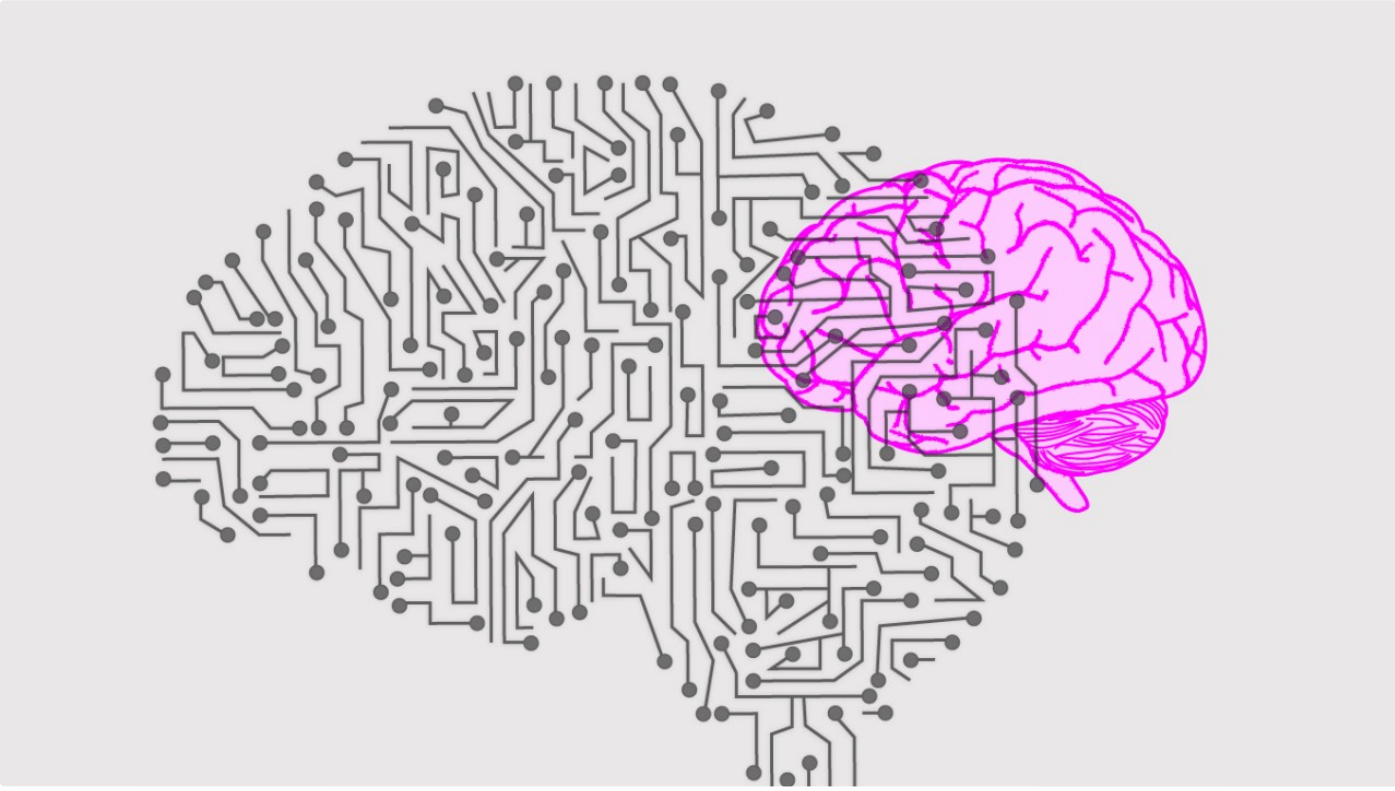
Dette indlæg er alene udtryk for skribentens egen holdning.

SYNSPUNKT

Fra AI til AGI i 2025: Hvad er det egentlig dagens AI mangler for at kunne blive til AGI, menneskelignende intelligens?

Kunstig Intelligens 26. marts kl. 05:30



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, som tidligere beskrevet i del 1 af dette indlæg her i DataTech. Her i del 2 fortsættes med at se lidt nærmere på, hvad det egentlig er dagens AI mangler for at bliver til AGI, menneskelignende intelligens. Simon Laub, lektor ved Erhvervsakademi Aarhus, går i dybden med AGI. Illustration: Ingeniøren.

Simon Laub
Lektor, Erhvervsakademi Aarhus

Der er ikke enighed om, hvad Artificial General Intelligence præcist er. Menneskelignende intelligens kunne være et bud.

Men mennesker er ikke lige intelligente på alle områder. Nogle mennesker er meget musikalske, nogle er gode til logisk tænkning, andre har stor selvindsigt, osv. AI har for længst overgået menneskelig intelligens indenfor specifikke områder, såsom at spille skak, hurtigt at kunne oversætte mellem alverdens sprog, forudsige 3D-strukturer i proteiner, m.v.

Men det er selvfølgelig noget ganske andet at have maskiner med generel intelligens, der kan udføre en bred vifte af opgaver på samme niveau som et menneske.

Og det er jo så her, AI-sprogmodellerne bliver interessante, da de jo tilsyneladende nu kan tale med om hvad som helst.

ChatGPTs skabere, OpenAI, fortæller os ikke præcist, hvad de arbejder med for at introducere mere menneskelignende ræsonnering i modellerne (i de såkaldte reasoning models), men det virker plausibelt, at det kunne handle om at dele udarbejdelsen af svar på svære problemer op i mindre skridt ('chain of thoughts), prøve forskellige ræsonnerings-'veje' og lave forskellige slags valideringer undervejs. Hvis en bruger, dvs. millioner af brugere, er tilfredse med et ræsonnement, kan systemet notere sig det, ligesom systemet også kan notere sig, hvis brugere er utilfredse. Og på den måde bliver bedre.

Det var således med en vis spænding, mange af os i efteråret sad klar til at teste modellen ChatGPT o1. Mange havde en idé om, at hvis man skulle finde frem til ting som modellerne var mindre gode til, så skulle man komme på en problemstilling, der hidtil var uset, og måske lave spørgsmål, hvor løsningen er mindre baseret på ord og sprog, og mere på f.eks. forestillingsevne.

Så i en populær problemstilling beder man ChatGPT o1 om at »forestille sig en kasse med 4 huller, hvori man kan stikke en hånd

»forestille sig en kasse med 4 huller, hvori man kan stikke en hånd ind. Inde i kassen er der ved hvert hul en kontakt, som kan være *off* eller *on*. Man bestemmer selv hvilke huller man vil stikke sine hænder ind i. I hver runde kan man sætte disse kontakter, som man vil, herefter roteres kassen rundt af en super intelligent robot og lander på en ny måde, hvor man ikke ved hvad man har rørt ved før. Hvor mange runder skal man bruge for være sikker på, at alle kontakter er enten *off* eller *on*?«.

Med et stykke papir og lidt tid kommer mange mennesker på en strategi for, hvad de ville gøre for at løse problemet.

Men. Gys. Ville ChatGPT finde ud af det? Stod vi her overfor Artificial General Intelligence, menneskelignende intelligens?

Da jeg forsøgte mig, fik jeg en lang udredning fra ChatGPT o1, fyldt med mange mellemregninger og til sidst et meget præcist svar, hvortil modellen havde tilføjet QED, Quod Erat Demonstrandum, hvorefter man så kunne give sig i kast med at undersøge om alt virkelig var i sin skønneste orden.

Svaret var forkert.

Hvilket ikke ville have overrasket AI-forskeren Subbarao Kambhampati fra Arizona State University, der, med en baggrund i automatiseret planlægning, i de senere år har været en konstant stemme til at minde alle om, at kritisk tænkning er så vigtigt nu som nogensinde før: Sprogmodeller kan ikke verificere deres egne svar. Deler man et svært problem op i mindre dele, skal alle led i

argumentationskæden være korrekte, før man kan være sikker på, at det endelige svar er korrekt. Og der findes ikke nogen eksterne universelle mekanismer, man bare kan kalde undervejs for at få afgjort om et udsagn er sandt eller falsk.

Større er ikke bedre

Kambhampati mener heller ikke, at det er en løsning bare at lave sprogmodellerne endnu større, trænet på endnu større data mængder, i håb om at modellerne så skulle udvikle 'emergent behaviours', i form af menneskelignende ræsonneringsevner. Ifølge Kambhampati har alt for mange været alt for ukritiske, når de har rapporteret om såkaldte 'emergent behaviours'. Vi skal huske på at internettet er kæmpestort, og mange gange får vi sikkert bare ting tilbage, som i en eller form allerede står derude, og derfor ikke er original tænkning. I det hele taget er Kambhampati ikke en stor fan af 'emergent behaviours'. Når man bygger et IT-system, ønsker man at vide præcist, hvad man har bygget.

Det er ikke i sig selv prisværdigt, hvis ens IT-system udviser adfærd man ikke kan forklare. Et AI-system, der skal stå for togdrift må ikke hallucinere på jobbet, og mennesker ville med garanti blive urolige, hvis fysiske robotter omkring dem pludseligt begyndte at lave helt uventede ting.

På vej mod AI-systemer med egne modeller af verden

Og Kambhampati er ikke den eneste, der er noget forbeholden i forhold til hvor langt man egentlig kan komme med store sprogmodeller. Ifølge AI-forsker og Chief AI Scientist ved Meta, Yann LeCun, så er vi stadig meget langt fra AI-systemer der kan ræsonnere og planlægge på et menneskelignende niveau. Ifølge LeCun forudsiger og forstår sprogmodeller verden (dvs. tekster) i én dimension, mens AI-modeller der arbejder med billeder og

video forstår verden i to dimensioner.

Men for rigtigt at forstå verden, og kunne tænke, må man kunne lave simulationer i tre dimensioner. Man må have en verdensmodel - en 'world model' - hvori man kan afprøve og simulere handlinger i verden, inden man rent faktisk gør noget. Med sådanne simulationer kan man forklare, hvorfor man lavede en bestemt handling, og man kan blive overrasket, når man ser noget, der er i modstrid med ens model af verden.

Der er selvfølgelig mange problemer, der skal løses, inden vi kommer så langt, at vi lever i en verden med robotter, der går rundt med deres egne world models. LeCun peger f.eks. på problemet med at dele en overordnet plan op i små delplaner. Jo, hvis man skal fra USA til Frankrig, skal man nok flyve, men man skal også have en plan for at komme til lufthavnen, og inden da skal man nok stå op af sengen i ens hjem i USA. Selv med alle Facebooks ressourcer til rådighed er det ikke så nemt at skrive et program, der kan løse den slags problemer, helt generelt.

General intelligence er bare ikke nemt. Det handler ikke bare om at have flere 'narrow intelligences', men også om at få disse integreret sammen i en helhed. Menneskelige hjerner er ikke kun udstyret med sprogforståelse og sproggenerering i hjernens Wernicke- og Brocaområder, men har forbindelser på kryds og tværs til andre områder, som f.eks. præfrontal cortex, som kan medvirke til planlægning og beslutningstagning.

Måske er Artificial General Intelligence slet ikke lige om hjørnet. Og mange udsagn fra marketingsafdelinger rundt omkring i verden vil sikkert med tiden blive set i det samme lys, som vi nu ser de noget optimistiske udmeldinger AI-forskerne Simon, Shaw og Newell kom med i 1957, da de introducerede verden til GPS (General Problem Solver), et computerprogram der dengang i 1957 lovede at være en universel problemløser.

Det sagt, så er det værd at nævne, at selvom Kambhampati er skeptisk i forhold til mange udsagn om store sprogmodeller, forsømmer han aldrig at rose alt det modellerne er gode til, som idégenerering og meget andet. LeCun er selvfølgelig helt på det rene med, at der er lang vej endnu, inden vi har AI-systemer, der kan lave simulationer i deres world models, og derved forstå verden på en menneskelignende måde. Men ifølge LeCun er der stadig veje åbne for en fremtidig verden, hvor alle mennesker har et større personale (AI-agenter udstyret med AGI) ansat til at hjælpe netop dem.

Ifølge LeCun er der heller ikke grund til bekymring om, hvor en sådan historie mon ender. De intelligente agenter kommer ikke fra den ene dag til den anden. Vi skal tage problemerne i forhold til, hvor langt vi reelt er kommet. Der skal lidt fremskridts tiltro til. Og der er tid til at arbejde for den gode version af fremtiden. Derudover vil vi mennesker jo gerne forstå hvad intelligens er - både maskinernes og vores egen. Der skal arbejdes for sagen, men ifølge LeCun er vejen fremad god nok.

Og når han mener det, kan vi andre vel også mene det?

Ovenstående indlæg er 2. del af Simon Laubs, lektor ved Erhvervsakademi Aarhus, serie om AI til AGI. Den første kan du læse eller genlæse her.

Vil du bidrage til debatten med et synspunkt? Så skriv til vores debatredaktion på debat@ing.dk

Denne artikel

Techgiganter til Trump: Giv os lov til at træne AI på rettighedsbeskyttet materiale

Fra AI til AGI i 2025: Tyder AI-udviklingen fra de seneste årtier på et stort gennembrud, eller gentager historien blot sig selv?

Emner

Kunstig Intelligens Følg Robotter Følg Softwareudvikling Følg

Data Følg

Omtalte virksomheder

Facebook, OpenAI