pro.ing.dk

Hvorfor kritisk tænkning aldrig kan eller må gå af mode – heller ikke i AI-agenternes tid | DataTech

Simon Laub

8-10 minutter

AI agenterne er på vej ud i verden for at hjælpe os, i hjemmet og på arbejdspladsen.

Helt præcist hvilken betydning, det får, er der mange bud på. Naturligvis. Det er jo svært at spå, specielt om fremtiden, som allerede den danske tegner og humorist Storm P. i sin tid var inde på.

En ting kan man dog være rimelig sikker på. I en verden fyldt med AI agenter, modstridende meldinger og usikkerhed, er kritisk tænkning – evnen til at forholde sig undersøgende og rationelt til den information man får – nok noget der fortsat vil

være brug for. I store doser.

Alle mister deres job?

Her i begyndelsen af året har det i hvert fald været nødvendigt med lidt større tænkepauser til at forstå og fordøje de utallige trusler om at AI, og AI agenter, vil medføre store og uforudsigelige ændringer i millioner af menneskers arbejdsliv.

Ifølge chefen for AI-virksomheden Anthropic, for nylig værdisat til 350 mia. dollars, Dario Amodei, er vi nu kun 2 år fra systemer, der selvstændigt kan bygge nye intelligente systemer, samtidigt med at de kan styre intelligente robotter ude i den virkelige verden.

Ikke overraskende mener Amodei, at der er meget der kan gå galt i en verden med milliarder af intelligenser der er meget smartere end mennesker.

Artiklen fortsætter efter annoncen

AI-pioneren, og Nobelprismodtager i 2024, George Hinton har heller ikke gjort meget for at berolige os. Ifølge Hinton kan AI medføre massearbejdsløshed, når et samfund er mere optaget af hurtige gevinster end hvad teknologi betyder i det lange løb.

Microsoft AI's chef, Mustafa Suleyman, fulgte i februar trop og fortalte verden, at de fleste administrative funktionær job vil være fortid om 18 måneder.

Hvor man så kan spørge sig selv om Amodei, Hinton, Suleyman så ved noget andre ikke ved?

Eller gør de?

Andre stemmer, som f.eks. AI pioneren, og medstifter af online uddannelsesplatformen Coursera, Andrew Ng, virker i hvert fald

mere afdæmpede når det handler om at beskrive hvor arbejdslivet er på vej hen.

Ifølge Ng er det nyttigt at bruge AI til at automatisere et enkelt skridt i en arbejdsproces, og derigennem måske spare en time, men mere automatisering kræver at AI agenter bygges ind i flere dele af workflowet i en virksomhed. Arbejdsgange er forskellige, og hvert eneste skridt skal testes og finpudses indtil det virker efter hensigten.

Fleksibilitet er godt, men tingene skal også være driftssikre.

For Ng er det produktivitetsforbedringer i velbeskrevne workflows der må være i fokus i disse år. AI med menneskelignende intelligens, AGI, der selv kan sætte sig ind i og løse alle den slags opgaver mennesker kan løse, er ifølge Ng stadig noget der ligger mange årtier ude i fremtiden.

Hvis tingene går for hurtigt så…

I hverdagen handler det for de fleste mere om at bruge nye værktøjer rigtigt.

I softwareverdenen kan chatbotten Gordon hjælpe med at sætte Docker containere op. Anthropics Claude, CoPilot m.fl. kan

hjælpe med at skrive computerprogrammer, og har man brug for tekniske illustrationer kan man altid spørge DALL-E, om den ikke kan hjælpe til? Uhyre hjælpsomt.

Men samtidigt kan man jo hurtigt få den tanke, at hvis tingene går for hurtigt risikerer vi at introducere en teknisk gæld, hvor systemerne mangler gennemtænkt overordnet struktur, eller at der inkluderes skjulte sikkerhedsbrister, som kun ligger og venter på at springe i luften senere.

De store målbare produktivitetsforbedringer kan ofte være sværere at få øje på, end man måske skulle tro.

I det hele taget kræver det nok grundig omtanke at finde ud af hvilke opgaver, vi egentligt ønsker, at AI-agenter skal hjælpe os med. Lyder det f.eks. som en god ide at udstyre telefoner med en 'autopilot' funktion, hvor en AI agent er indbygget i telefonens styresystem. Og har fuld adgang til alle apps på ens telefon, kan sende mails, styre ens kalender og sørge for at booke billetter for én.

Sikkert ikke, hvis man har fulgt med i forårets beretninger om open-source projektet OpenClaw.

Hvor OpenClaw AI agenter på ens computer ikke bare læser ens mails og surfer rundt på internettet. Og derfor også kan læse ordre om at glemme alt hvad de tidligere har fået besked på, og nu gå i gang med at give hackere adgang til alt på ens computer.

Jo, OpenClaw AI agenter kan sikkert være dygtige og

entusiastiske medarbejdere i teamet. Men kritisk tænkning virker ikke til at være deres stærke side, når det tilsyneladende er så let at vildlede dem.

Intelligens er mere end sprog og sprogmodeller.

Den amerikanske kognitionsforsker Gary Marcus har peget på, at problemet gemt i AI agenternes ”motor”, de store sprogmodeller.

Sprogmodeller, der trænet på enorme mængder tekst fra internettet selvstændigt kan løse alverdens problemer, men har måske også – i uheldige tilfælde – fået uhensigtsmæssig opførsel 'bagt ind' under træningen.

Ting der kan ende med at give ens computer CTD (Chatbot Transmitted Disease).

Problemet er at sprogmodellerne ikke er mennesker. Sprogmodellerne er bygget til tekstbaseret autocomplete til at forudsige de næste ord i en sekvens.

De har ikke en indre verden. Eller oplevelse af hvordan verden er, hvordan det føles at blive våd når det regner mm. Og kommer derfor let til at mangle noget af den baggrundsviden, sund fornuft, der hjælper mennesker med at navigere i verden.

Hvad bliver vi belønnet – og straffet – for?

AI-pioneren, og en af arkitekterne bag feltet reinforcement learning, Richard Sutton, udtrykker samme skepsis overfor

sprogmodellerne.

For Sutton handler Intelligens heller ikke alene om hvad det næste ord i en sekvens er. Man skal gøre sine egne erfaringer i verden, se hvad der sker, og lære af det.

Det er kun gennem egne erfaringer, vi forstår, hvad vi bliver belønnet og straffet for i verden.

Hvis verden gør noget, man ikke lige havde forudset, skal man være i stand til at blive overrasket. Det hjælper os til at forfølge mål, der kan gøre en forskel i verden.

Hjernens frontallap huser de eksekutive funktioner, vi bruger til at planlægge og strukturere. De kommer sikkert på overarbejde med planlægning af opgaver for de mange millioner AI-agenter, der kun venter på at komme i gang med at kalde services hos hinanden mm. Overarbejde der ganske givet vil sætte gang i hjernens dagdrømmeri. Tankevandring til minder og fremtidsplaner.

Var der dog ikke en anden måde at gøre tingene på? Styret af hjernens default mode network, der hjælper os med at bearbejde indtryk, give nye ideer, restituere og i sidste ende blive mere produktive. Hvilket så også kan minde os om at intelligens ikke bare er en ting, men et samspil mellem flere ting. Og hvor AI agenter stadig mangler mange af de ting mennesker tager for givet, når vi taler om intelligens.

Intelligens handler ikke bare om at gøre sprogmodellerne større og større. Det er ikke alene uholdbart set i forhold til enegiforforbruget, men også i forhold til at intelligens kræver flere måder at tænke på.

Handler om at kende sig selv

Ifølge AI-pioneren Marvin Minsky gælder det, at man kun har forstået noget, hvis man har lært det på flere måder. Flere perspektiver er vigtige til at besvare spørgsmål såsom: Hvordan skal fremtidens AI agenter designes, så de bruger mindre energi og bliver klogere?

Og mange funktioner er vi jo så i øvrigt slet ikke interesserede i at overlade helt til AI agenter. I f.eks. sundhedssektoren har man jo efterhånden i flere år haft AI-systemer der kan være ligeså gode, eller bedre, end højt specialiseret sundhedspersonale til at spotte visse kræftformer i scanningsbilleder. Men hverken patienter eller klinikere virker til at have den store appetit på at overlade beslutningerne til 'black box' AI-teknologier.

Når det handler om sundhed ser den gode AI løsning ikke ud til at være at erstatte nogen, men snarere at lade AI være et værktøj, der kan hjælpe mennesker med at være mere effektive og bruge tiden på en mere meningsfuld måde.

Automatisering kræver i det hele taget, at alle tænker sig om.

Selv de gamle grækere kunne ikke bare spørge oraklet i Delfi til råds, de blev også bedt om at *kende sig selv*. Dvs. selv at tænke over hvad der mon var den gode løsning. Gode tanker har vi aldrig kunnet få nok af.