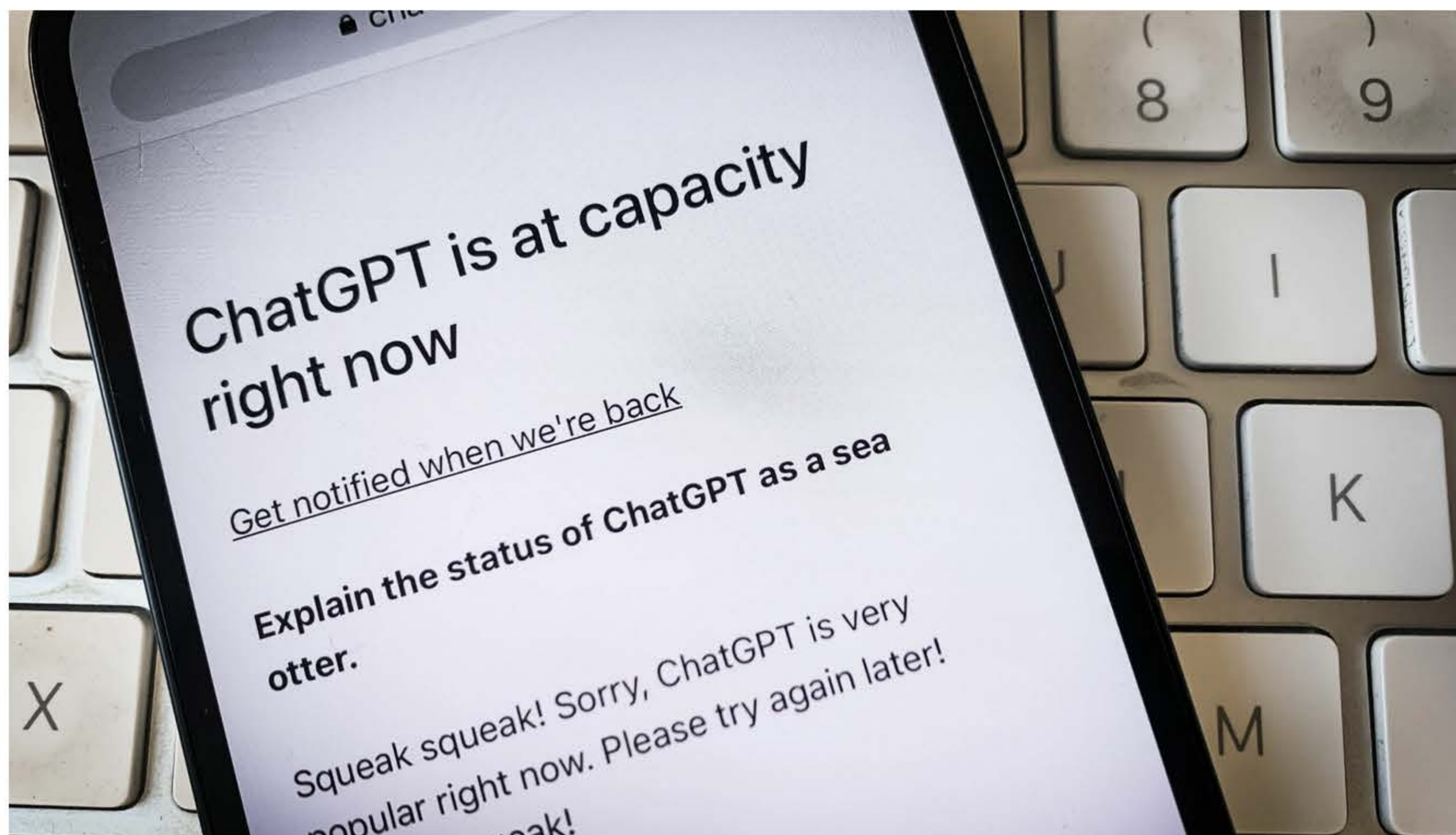


DEBATT • KETIL STØLEN

I hvilken grad vil maskinlæring endre forskningen?

Større datasett, mer praktisk hjelp og effektivisering, mener Ketil Stølen.



En støtte og en samarbeidspartner for framtidens forskere. Det mener iallfall artikkelforfatteren. Foto: Espen Løkeland-Stai



Sjef forsker ved Sintef og professor II ved UiO
Ketil Stølen

Publisert Mandag 17.04.2023 - 10:42



Maskinlæring (eller ML) er i vinden som aldri før, ikke minst teknologier basert på store språkmodeller som Chat-GPT. Det hevdes at ML endrer måten vi forsker på, at forskerens rolle forandres og at det i fremtiden trengs færre forskere. I hvilken grad er dette riktig? I det følgende skal vi se nærmere på disse påstandene.

I fremtiden vil «alle» bruke ML, forskerne inkludert. Noen forskere vil kun anvende eksisterende ML-modeller, som ChatGPT er et eksempel på. Andre vil lage nye ML-modeller og mange vil gjøre begge deler.

I boka «Argumentasjonsteori og Vitenskapsfilosofi» fra 1977 sier Føllesdal og Walløe følgende: «Et viktig aspekt ved forskningsprosessen er at arbeidet som regel (kanskje alltid?) foregår i forhold til hypoteser. I noen tilfeller er relasjonene til hypoteser lette å få øye på, og forskeren er selv klar over at han forsøker å finne støtte for noen hypoteser, at han forsøker å avkrefte andre hypoteser, eller at arbeidet nettopp består i å sette frem nye hypoteser (som kanskje kan forklare fenomener som ikke så lett lot seg forklare fra de gamle hypotesene). I andre tilfeller kan hypotesene være vanskeligere å få øye på.»



DEBATT • HANSEN OG VESTBO
Ja til KI-pause

ML endrer ikke dette. En ML-modell er fremstilt av et dataprogram som har utført en læringsprosess. Menneskelig design av ML innebærer å designe denne læringsprosessen. En forsker designer læringsprosessen med utgangspunkt i en problemstilling. Dette innebærer å velge hvilken læringsmetode som skal benyttes, ekstrahere datasett i form av treningsdata og testedata, definere belønningsfunksjoner, osv. Treningsdataene benyttes til opplæring, mens testedataene brukes til å sjekke om læringen var vellykket. Læringsprosessen kan finne sted offline, online, eller begge deler. Det er altså ikke slik at dataene alltid hentes inn på forhånd.

At det valgte læringsprosessdesignet er egnet for den aktuelle problemstillingen, er en slags implisitt arbeidshypotese. Hvis designet er bra, er resultatet av å utføre læringsprosessen en ML-modell som løser problemet.



FUSK

Lot ChatGPT skrive artikkel om juks. Lurte fagfeller trill rundt

Litt forenklet kan vi tolket utsagnet «ML-modellen løser det aktuelle problemet» som en hypotese om den genererte programvaren. Denne hypotesen kan være feil i likhet med enhver annen hypotese. Dette vil ofte oppdages ved hjelp av testedataene. Men selv om ML-modellen overlever testingen kan hypotesen likevel være feil. Dataene som læringen baserte seg på kan for eksempel ha vært utilstrekkelige, dårlig valgt eller manipulert, og programvaren kan lide av alle slags svakheter. Derfor krever denne hypotesen den samme grundige evalueringen som enhver annen forskningshypotese. Igjen er ML et nyttig verktøy. ML kan også identifisere sammenhenger og mønstre i dataene som det ellers hadde vært vanskelig å oppdage, og i så måte bidra med nye hypoteser.

Men hva hvis dataene er kvalitative, for eksempel resultater fra intervjuer utført i henhold til en intervjuguide? I utgangspunktet har man kanskje ingen hypotese om hva forskningen vil resultere i, men man har i det minste en implisitt metodeorientert hypotese om at ved å utføre forskningen i henhold til den valgte metoden vil man avdekke noe om virkeligheten.



DEBATT • HELGE HØIVIK
(Nok) Et slag for PISA-modellen

Tradisjonelt har denne type kvalitativ forskning vært svært arbeidskrevende. Intervjuene må for eksempel renskrives, kodes og deretter analyseres. Store språkmodeller vil effektivisere dette arbeidet. De kan utføre slike analyser både bedre og raskere enn mennesker, og de vil til og med kunne utføre intervjuer. Forskernes tid kan dermed frigjøres til design av forskningsmetode og overordnet styring av prosessen. Likevel vil dette neppe redusere antallet «kvalitative forskere». For det første: store språkmodeller vil muliggjøre langt større studier med vesentlig høyere validitet. I stedet for å intervju syv personer kan man nå intervju 70. Forskerens oppgave blir å planlegge, styre gjennomføringen og kvalitetssikre det hele. For det andre: ML vil med stor sannsynlighet øke behovet for kvalitativ forskning, ikke minst fordi teknologien selv vil gi opphav til store samfunnsmessige endringer som vi er helt avhengig av å forstå.

ML-modeller er nyttige verktøy som dyktige forskere vil arbeide i to-spenn med. ML-modeller vil ta seg av det arbeidskrevende som å oppsummere andres arbeid, analysere data, oversette, gi råd og teste. ML-modeller vil effektivisere forskningen og muliggjøre forskning som inntil nylig har vært umulig i praksis.

Rediger

! Denne teksten er et debattinnlegg. Innholdet i teksten uttrykker forfatterens egen mening.

LES VIDERE ETTER ANNONSEN

FÅ NYHETER PÅ MOBILEN

Last ned Khrono-appen!

Last ned fra
App Store

LAST NED PÅ
Google Play