

## ChatGPT – en høflig og gudfryktig samtalepartner som alltid taler sant, eller?

2022 var året da kunstig intelligens virkelig inntok den offentlige samtalen i Norge, og i andre deler av verden. De siste ukene har debatten i Norge gått om hva kunstig intelligens kan bety for juks og [læring](#) i [skoler](#) og [universiteter](#), for [jobben til kunnskapsprodusenter](#) – men også for [framtida til en stor søkemotor som Google](#). Kort sagt, hva blir konsekvensene av en ny og potensielt disruptiv teknologi.

Utgaven ChatGPT har stått sentralt i den norske debatten på tampen av 2022, og det viser seg at den kan langt mer enn kommune-Kari. Det interessante med den offentlige samtalen knyttet til ChatGPT, er at premisset synes å være dette: [Denne roboten er mer tidsbesparende for deg enn en søkemotor, fordi ChatGPT oppsummerer det som står i kildene for deg](#). Oppsummert i klart og saklig språk kan den på få sekunder gi deg korrekt informasjon om hva som helst.

Hva om dette ikke stemmer? Hvordan kartlegge dette? Hva med å spørre den selv?

Chat GPT er ærlig om noen av sine begrensinger. På spørsmål om seertall ved årets World Cup (sport ikke oppgitt) svarte den at den ikke hadde anledning til å søke på nettet, og at den ikke hadde nyere info tilgjengelig enn fra 2021. (Men den skjønnte straks at det dreide seg om fotball!)

Så til alle lærere som trenger et rødt flagg for juks: La deler av oppgaven bli knyttet til så nye hendelser som mulig.

Chat GPT er heller ikke laget for å tegne, regne eller forklare mindre kjente konsepter innen fysikk - f.eks. absisjon/absement (tidsintegralet av posisjon, kan bl.a. brukes som mål på hvor mye vann som har kommet ut av en kran man har åpnet i varierende grad i varierende tid). Hvis den blir utfordret på disse områdene gjør den noen ganger et forsøk som heller tvilsomt ville gitt ståkarakter. Andre ganger innrømmer den at dette ligger utenfor kompetanseområdet, som tross alt er språk. (Den klarte likevel å løse oppgaver ganske langt utenfor kjernekompetansen - som å gi verdier for spesifikke verdier matet til ikke-elementære matematiske funksjoner.)

Ellers skal det sies at ChatGPT har med seg noen mer normative elementer også. Den har et høflig språk. Den takker for komplimenter – og for konstruktiv kritikk. Tydeligvis er den også programmert til, eller den har erfart seg fram til, at den skal holde seg nøytral til potensielt polariserende tema som religion. Den har imidlertid erklært seg som ikke-ateist, sågar hevdet (kanskje på spøk) at [«Gud har gitt oss AI-teknologien som eit verktøy til å betre academia.»](#)

[Noen spekulerte på om kunstig intelligens faktisk er ærligere enn mennesker](#), og viser til eksempler på bot-er som er [spesialister på spill](#). Kan det tenkes at bot-en bare gjorde det nestbest i et strategispill som *Diplomacy* fordi den ikke dolket noen i ryggen? Tja, akkurat her er det kanskje mer interessant å skjele til hva forskerne bak forsøket sier dette: [AI ble nemlig straffet for kontraktsbrudd](#).

Og så til den store testen i vårt forsøk med ChatGPT: Hva kan den om latin, det beryktet vanskeligste språkfaget siden Alexander Kiellands dager. Joda, bot-en har latin på repertoaret. Den kan generere eksempelsetninger på latin, der den selv finner på hva de skal bety. Så langt, så vel.

Kan den også hjelpe oss med å bli kjent med den virkelig gamle litteraturen på latin? Den fikk spørsmål om å gjengi Ovid Amores 3.2, og kom kjapt opp med dette:

Quis tibi, mea Lesbia, quis mille dolit  
carmine tales, quae legis, ferre querelas?  
qui facis, ut totiens facias puella fateri  
se tibi semper parere et debere merenti?

*Som er helt feil.*

Da den så ble konfrontert med muligheten for at dette kanskje ikke stemte, men en såpass snill formulering som "Are you sure that is actually the text of it? That sounds more like Catullus in my ears." svarte den med dette:

I apologize for the mistake. The text you provided is actually from Catullus Poem 5, not from Ovid's Amores. Here is the text of Ovid's Amores 3.2:

At mihi non licuit tota cum duplice mente  
vivere, et duobus oculis te, dea, videre.  
quae mihi diva tuos incidere luctus in antros  
fecisti, ut sola inter speluncas ego tenebrae  
tenerer, et tua me somnia fallentia tangent  
semper, et per noctem nigram virgo pudica vaget.

Feil igjen. Dette er heller ikke Ovids Amores 3.2 og det første diktet er ikke Catullus 5, *faktisk synes ingen av disse to diktene å være tidligere eksisterende latinske dikt i det hele tatt* (utfra det vi hadde tilgang på av andre kilder). Vi ble her nysgjerrige på om den ville vise tilbake til det første diktet om vi nå spurte om Catullus 5, men resultatet ble da:

Lesbia, mea Lesbia, si tibi di lecti  
eheu quid faciunt homines? quid faciunt? quid faciunt?  
nuper me in lupanari, pueri, vidistis,  
cum meretriculae viridantes oscula darent,  
cinaede et optatae formosae parataeque.

Dette er feil, for tredje gang. Da ChatGPT ble konfrontert med at den nå hadde påstått at to forskjellige dikt var Catullus 5 presterte den omsider å finne et riktig dikt:

Vivamus, mea Lesbia, atque amemus,  
rumoresque senum severiorum  
omnes unius aestimemus assis.  
soles occidere et redire possunt:  
nobis cum semel occidit brevis lux,  
nox est perpetua una dormienda.

Dette åpner et nytt landskap knyttet til spørsmålet om påliteligheten til kunstig intelligens. Kildekritikken som er nødvendig å bruke på et søkemotor-funn gjelder i samme grad for det ChatGPT serverer, så pass på å be den vise deg hvilke kilder den bygger på (og finn ut hva den svarer da!). Og kanskje ligger det noen dypere spørsmål om pålitelighet her også, som vi kan glemme litt bort i vår iver etter å omfavne ny teknologi. Når skal vi stille disse spørsmålene, og til hvem?

Vårt lille forsøk tyder på at ChatGPT rett og slett bløffet. Fortsatt like keen på å la denne skrive eksamen for deg?