
Kan kunstig intelligens være empatisk?

INVITERT KOMMENTAR

NIKOLAI OLAVI CZAJKOWSKI

n.o.czajkowski@psykologi.uio.no

Nikolai Olavi Czajkowski er førsteamanuensis ved Promenta forskningscenter, Psykologisk institutt, Universitetet i Oslo og forsker ved Folkehelseinstituttet.

Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Når helsepersonell vurderer svar på spørsmål om sykdom og symptomer, besvart av enten leger eller kunstig intelligens, ser de ut til å foretrekke svar fra sistnevnte.

Mork og medforfattere tar utgangspunkt i 192 helserelaterte spørsmål fra nettstedet Studenterspør.no i studien som nå publiseres i Tidsskriftet (1). GPT-4 ble brukt til å generere nye svar på spørsmål som allerede var besvart av leger. 344 respondenter med helsefaglig bakgrunn vurderte deretter svarene ut fra tre kriterier: empati, kunnskap og hjelpsomhet, uten å vite hvilken kilde svarene kom fra. Resultatene viste at svarene fra GPT-4 ble vurdert som mer kunnskapsrike, hjelpsomme og empatiske enn de som var skrevet av legene.

Mange fikk øynene opp for kunstig intelligens (KI) gjennom lanseringen av ChatGPT. Slike store språkmodeller er et eksempel på generativ KI, en teknologi som ikke bare kan klassifisere og predikere, men også skape nye data basert på eksempler den er trent på. Denne teknologien omgir oss med tekst, bilder, lyd og video som blir stadig vanskeligere å skille fra innhold produsert av mennesker. At generativ KI var på vei og var forventet å påvirke samfunnet bredt, lå til grunn for utarbeidelsen av vår nasjonale strategi for kunstig intelligens, som ble lagt frem i januar 2020 (2). Du skal lete lenge etter en mer varslet teknologisk revolusjon. Likevel har selv flere av utviklerne blitt overrasket over hvor raskt teknologien har utviklet seg, drevet av en voldsom økning i regnekraft og tilgjengelig treningsdata.

«Du skal lete lenge etter en mer varslet teknologisk revolusjon. Likevel har selv flere av utviklerne blitt overrasket over hvor raskt teknologien har utviklet seg»

ChatGPTs arkitektur og hvilke data som inngår i treningen, er ikke offentlig kjent, men vi vet at den er trent på en stor andel av all tekst produsert av mennesker. At modellen er kunnskapsrik, er derfor ikke særlig uventet, og den har allerede vist seg å prestere på et høyt nivå innen fagfelt som biologi og juss (3). Det som derimot kan være mer oppsiktsvekkende, er at slike modeller også evner å gi svar som oppleves som hjelpsomme og, ikke minst, empatiske.

Dette er også en viktig side av den revolusjonen vi står i. KI har vist seg å være fremragende, ikke bare i logikkbaserte strategispill som sjakk og Go, men også i å navigere komplekse sosiale situasjoner (4). Den har demonstrert en imponerende evne til å forstå andres perspektiver, ofte bedre enn de fleste mennesker (5, 6). Dette kan forklare hvorfor GPT-4 i studien til Mork og medforfattere oppleves som empatisk i møte med folks bekymringer. Samtidig er det viktig å huske at ChatGPT ikke føler noen reell empati, selv om det kan oppleves slik. Hvordan denne tilsynelatende empatien vil påvirke pasienter, er et åpent spørsmål. Kan den forbedre kommunikasjonen, eller risikerer vi at den skaper en situasjon der grensen mellom simulert og ekte omsorg blir uklart?

«At ChatGPT oppleves som informativ, hjelpsom og empatisk innen helseveiledning og rådgivning, betyr på ingen måte at teknologien er klar til å erstatte helsepersonell i pasientbehandling»

Med kraften til disse modellene følger betydelige etiske utfordringer. KI-modeller som ChatGPT er kjent for å «hallusinere», å generere feilaktig informasjon som ofte presenteres i en så flytende og veltalende form at den fremstår svært overbevisende. Dette er særlig problematisk innen medisin, hvor feilinformasjon raskt kan få alvorlige konsekvenser. At ChatGPT oppleves som informativ, hjelpsom og empatisk innen helseveiledning og rådgivning, betyr derfor på ingen måte at teknologien er klar til å erstatte helsepersonell i pasientbehandling. Her er Mork og medforfattere også tydelige: studien deres undersøker hvordan generativ KI kan bidra som et rådgivende verktøy, ikke som en erstatning for medisinsk ekspertise.

For hvem har ansvaret dersom KI gir feil informasjon? Hvordan sikrer vi at pasienter beholder tilliten til helsepersonell når de vet at svarene kan være KI-genererte? Og hvordan unngår vi at KI forsterker eksisterende skjevheter i helseinformasjon eller tilgang til helsetjenester? At ChatGPTs treningsdata holdes hemmelig, gjør det vanskelig å forstå hva modellen baserer sine svar på, og i hvilken grad den er påvirket av skjevheter i sine kilder.

KI kan ikke holdes ansvarlig for sine anbefalinger, og ansvaret for medisinske råd og beslutninger må fortsatt ligge hos helsepersonell. Likevel føyer studien til Mork og kolleger seg til en økende mengde forskning som indikerer at slike modeller kan fungere som verdifulle støtteverktøy for helsepersonell, og gi reell

avlastning i møte med økende krav til effektivisering. Med ansvarlig bruk og en tydelig forankring i pasientsikkerhet og etiske prinsipper, kan denne teknologien bli et viktig bidrag til en trygg og effektiv helsetjeneste.

REFERENCES

1. Mork ET, Mjøs HG, Jammer I et al. Kunstig intelligens og legers svar på helsepørsmål. Tidsskr Nor Legeforen 2025; 145. doi: 10.4045/tidsskr.24.0402. [CrossRef]
2. Regjeringen. Nasjonal strategi for kunstig intelligens. <https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/> Lest 19.1.2025.
3. Achiam J, Adler S, Agarwal S et al. Gpt-4 technical report. arXivorg. Preprint 15.3.2023. <https://arxiv.org/abs/2303.08774> Lest 21.1.2025.
4. Bakhtin A, Brown N, Dinan E et al. Human-level play in the game of Diplomacy by combining language models with strategic reasoning. Science 2022; 378: 1067–74. [PubMed][CrossRef]
5. Kosinski M. Evaluating large language models in theory of mind tasks. Proc Natl Acad Sci U S A 2024; 121: e2405460121. [PubMed][CrossRef]
6. Strachan JWA, Albergó D, Borghini G et al. Testing theory of mind in large language models and humans. Nat Hum Behav 2024; 8: 1285–95. [PubMed][CrossRef]

Publisert: 10. februar 2025. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.25.0044
Opphavsrett: © Tidsskriftet 2026 Lastet ned fra tidsskriftet.no 23. juni 2026.