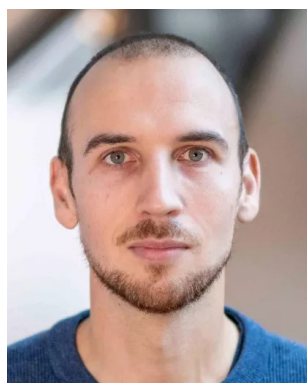


KI får ikke dårlig samvittighet



Magnus Helgheim Blystad
Oslo Nye Høyskole
magnus.blystad@oslonh.no

For å kunne utnytte mulighetsrommet i KI når det gjelder psykisk helse, må vi ha KI-ferdigheter for å kjempe mot utfordringene.



Magnus Helgheim Blystad

Foto: Oslo Nye Høyskole

Mulighetsrommet til kunstig intelligens (KI) er stort, også når det gjelder å bedre vår psykiske helse. Det er for meg åpenbart at KI, som da kan være en samtalepartner som har en støttende holdning, og tilgang til store deler av internett, kan hjelpe. Det har vært flere saker der KI har vært til hjelp for mennesker i vanskelige livssituasjoner, som ved depresjon og avhengighet. Et enkelt nettsøk gir deg mange forumposter om at KI har fungert som terapeut. Det er vanskelig å si til disse menneskene at det var feil å bruke verktøyene som hjalp dem. For mange er terskelen for å henvende seg til helsevesenet svært høy, både økonomisk og sosialt.

Jeg mener likevel det i dagens samfunn er for lite snakk om de potensielt negative sidene av KI. Nyanseringen må på plass, og det er dette som ofte savnes i offentlig debatt om KI. Mulighetsrommet KI tilbyr, må vi benytte oss av, men ikke ukritisk. I teksten vil forkortelsen KI brukes som en forenkling for chatboter drevet av store språkmodeller.

Drøm eller mareritt?

I det følgende vil jeg trekke fram eksempler der KI kan ha bidratt negativt til den psykiske helsen. En sak omhandler en mann som tok sitt eget liv etter at en KI oppfordret ham til å ofre seg for klimaet (Euronews, 2023). Året etter kom en ny selvmordssak (NBC News, 2024), og i år enda en (BBC News, 2025a). Det er lite forskning på hvordan KI håndterer temaet selvmord. Men det ser tilsynelatende ut til at KI-ene *kan* svare brukere på måter som ikke nødvendigvis bidrar til at de søker eller får hjelp (McBain et al., 2025). Fenomener som *KI-psykose* har også begynt å dukke opp, en tilstand hvor mennesker forsvinner hen til en forvrengt verden, geiledet av KI (BBC News, 2025b).

I de ovennevnte sakene finner man personlige tragedier hvor det er åpenbart at KI har spilt en rolle. Spørsmålet er hvilken? Og viktigere, hvordan kan vi forsvare oss mot det? Jeg mener

utfordringene med KI bør være en del av den offentlige debatten. Spesielt på samfunnsplan, når vi diskuterer tilgjengelighet og regulering av teknologien.



Språkmodeller får ikke skyldfølelse

Jeg mener det første vi må huske på, er at språkmodellene ikke kan forme en gjensidig relasjon med et menneske. I egen forskning har jeg gått inn i begrepet empati (se f.eks. Blystad & Grøgaard, 2024) og fattet stor interesse for studier som viste at utsagn fra en KI ble rangert som like empatisk som de fra mennesker (Czajkowski, 2025). Enkelte KI-boter kan komme med medisinske uttalelser som både var faglig korrekte og opplevd formidlet på en mer empatisk måte enn fra klinikere (Chen et al., 2025). Som forsker er det vanskelig å ikke bli fascinert. Men uansett bør vi nok først diskutere hva empati faktisk er, og om det som utspiller seg mellom et menneske og en KI-terapeut, kvalifiserer til betegnelsen.

Det er lett å bli ledet hen til at slike språkmodeller har en form for personlighet, eller til og med bevissthet. Flere brukere har havnet i eksistensielle samtaler med bruk av mystiske og gjentakende symboler. Diskusjonen om bevissthet går jeg ikke inn i. Men jeg mener det er mulig å anse slike modeller, for enkelthets skyld, som en type kalkulatorer. Gjør man det, blir følgende lett å forstå: Linjene med kode vil aldri føle på ansvar eller skyld. De vil aldri motiveres av et ønske om å være et prososialt tilskudd til samfunnet. Men i motsetning til kalkulatoren så kan vi lett *forvente* det, særskilt når de bidrar til at mennesker med psykiske vansker får det bedre.

AI literacy – KI-ferdigheter

Før var chatboter noe man møtte i serviceøymed. De ga automatiske svar på spørsmål som å opprette en bankkonto eller returnere en vare. Med KI får man unike responser som kan virke mer menneskelige. Men responsene er formet av matematikk og trening på datasett. Ingen menneskelig egenskaper er involvert.

Generative KI-er finnes allerede i alle apper. Alle barn som vokser opp i en verden av sosiale medier, vokser også opp i en verden med generativ kunstig intelligens. De som starter på skolen nå, vil ikke huske en verden uten. Det er på høy tid at vi finner ut hvordan det store fellesskapet i samfunnet kan komme på et kompetansenivå som lar oss bruke disse såkalte kalkulatorene på en vettug måte. Det handler i liten grad om å forstå matematikk og logikk, men konsekvenser av samhandlingene.

Vi har enda lite forskning, men tragediene beskrevet ovenfor understreker behovet for å kartlegge hva slags ferdigheter vi trenger for å forebygge mulige uheldige konsekvenser. Som en start mener jeg vi trenger å være bevisst særlig to ting: En KI har ingen menneskelig intensjon, samvittighet eller ansvarlighet, og responser fra KI er basert på trening på et datasett. Disse tingene vil nok i større grad være en del av diskursen om vettug bruk av KI, som et slags naturlig utgangspunkt. For vi må ta i bruk mulighetene, både i terapiøymed, læring og sikkert mye annet, men vi må bruke det på en kritisk og overveid måte.

Referanser

- BBC News. (2025a, 12. mai). Parents of teenager who took his own life sue OpenAI. *BBC*. <https://www.bbc.com/news/articles/cgerwp7rdlvo>
- BBC News. (2025b, 2. juni). "AI psychosis": People convinced imaginary things are real after AI chatbot use. *BBC*. <https://www.bbc.com/news/articles/c24zdel5j18o>

- Blystad, M. H. & Grøgaard, S. C. (2024). Empatiens mange ansikter: Fenomenologisk og positivistisk empati. *Tidsskrift for Norsk psykologforening*, 61(1), 32–43. <https://doi.org/10.52734/ERKP8472>
- Chen, D., Chauhan, K., Parsa, R., Liu, Z. A., Liu, F.-F., Mak, E., Eng, L., Hannon, B. L., Croke, J., Hope, A., Fallah-Rad, N., Wong, P. & Raman, S. (2025, 13. mai). Patient perceptions of empathy in physician and artificial intelligence chatbot responses to patient questions about cancer. *npj Digital Medicine*, 8, Article 275. <https://doi.org/10.1038/s41746-025-01671-6>
- Czajkowski, N. O. (2025). Can AI be empathetic? *Tidsskrift for Den norske legeforening*, 145. <https://doi.org/10.4045/tidsskr.25.0044>
- Euronews. (2023, 31. mars). Man ends his life after an AI chatbot 'encouraged' him to sacrifice himself to stop climate change. *Euronews*. <https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate->
- McBain, R. K., Pedersen, E. R., Bravata, D. M., Sontag-Padilla, L., Jaycox, L. H. & Stein, B. D. (2025). Evaluation of alignment between large language models and expert clinicians in suicide risk assessment. *Psychiatric Services*, 76(9), 875–882. <https://doi.org/10.1176/appi.ps.20250086>
- NBC News. (2024, 14. april). Lawsuit claims Character. AI is responsible for teen's suicide. *NBC News*. <https://www.nbcnews.com/tech/characterai-lawsuit-florida-teen-death-rcna176791>

