

Kunstig intelligens i terapirommet



Torhild Tønnesen Haugnes

Modum Bad

TorhildTonnesen.Haugnes@modum-bad.no

Roland Westerlund

Modum Bad

Forfatterne har fylt ut interessekonfliktskjema og oppgir ingen interessekonflikter.

Etisk bruk av KI krever klare ansvarslinjer med psykologen i førersetet.



Illustrasjon: Kristian Utrimark

Ved Modum Bads forskningsinstitutt arbeider vi med flere forskningsprosjekter som omhandler kunstig intelligens i terapirommet. Kunstig intelligens (KI) brukes i økende grad i psykoterapi. Teknologien finnes både i systemer som bistår klinikere med administrative oppgaver som oppsummeringer og journalnotater, og som beslutningsstøtte i diagnostikk og i prediksjon av behandlingsforløp ved hjelp av algoritmer. KI ligger også til grunn for interaktive terapeut-chatboter og pasient-chatboter, der sistnevnte kan hjelpe terapeuter under opplæring med å utvikle ferdighetene sine.

I teksten her ser vi på hvilke etiske implikasjoner KI kan få for pasientforløpet, med særlig vekt på bruken av KI som støtte for kliniske vurderinger. Vi ser spesielt på KI som aktør når det gjelder prediksjonsalgoritmer, en voksende trend innen beslutningsstøtte og persontilpasset behandling. Hvordan kan vi som klinikere møte de etiske implikasjonene prediksjonsverktøyer medfører rundt

spørsmål om tillit, transparens og mulighetene for svekket ansvarsfølelse. Vi spør om KI er en løsning, eller enda en kokk som lager søl i det intrikate samspillet ved et sykehus. Vi spør også hva de etiske implikasjonene er i lyset av at det emosjonelle ubehaget som følger av vanskelige prioriteringer, trolig aldri erfares av en algoritme.



Et pasientmøte i nær fremtid

La oss skru tiden noen år fram og forestille oss hvordan en klinisk hverdag kan se ut når dagens forskningsprosjekter er i bruk. På overflaten er mye det samme. En forskjell er at informasjonen som pasienten gir oss, behandles av algoritmer med potensial til å forutsi behandlingsforløp med høyere treffsikkerhet enn ved tverrfaglige kliniske vurderinger alene. Algoritmene ser mønstre som klinikerne kanskje ikke selv fanger i et hav av informasjon. Algoritmen har nå en kalkulert vurdering om hvilken behandling som er mest egnet til pasienten, og om pasienten er på vei mot et godt behandlingsutbytte.

Maskinlæring – den metodiske grunnpilaren for KI – vil etter hvert kunne bidra med prioriteringsstøtte i psykisk helsevern. Flere forskningsmiljøer har i dag kommet langt i å utvikle sofistikerte algoritmer som viser lovende resultater (se f.eks. Rubel & Lutz, 2025).

Ved Modum Bad forsker vi for å se om KIs algoritmer har bedre predikasjonsverdi i konkrete spørsmål som *hvilke* pasienter bør prioriteres, og *hvilke* behandlingsformer som er mest hensiktsmessige for den enkelte. Arbeidet vil også undersøke om algoritmer forutsier pasientens behandlingsrespons.

Hvis evidens underbygger økt bruk av algoritmer, kommer det en ny aktør inn i samspillet mellom pasientens, klinikerens, klinikkens og samfunnets interesser.

Ansvar og empati i prioriteringsarbeid

Prioriteringer er – på godt og vondt – en kjernekompetanse for klinikere. I Tidsskrift for Norsk psykologforening etterlyste Elvik (2025) bedre vurderingskompetanse og vilje til prioriteringer blant psykologer. Mange pasienter bærer med seg vonde avslagshistorier, politikere ønsker lavere avslagsprosent samtidig som utgiftene skal holdes nede, og klinikere må balansere inklusjonskriterier, offentlige veiledninger og klinisk skjønn. Prioriteringsarbeid er både personlig, politisk og fagetisk. Hva skjer når algoritmer og KI trer inn i dette allerede betente landskapet?

Pasienten deler informasjon i forventning om at vurderingene fattes på riktig grunnlag. KI viser lovende takter når det gjelder å redusere bias i kliniske vurderinger. Men dersom selvrapportskjemaer brukes som grunnlag for prediksjoner, oppstår et behov for å forstå *hvorfor* algoritmen anbefaler en bestemt behandling. Hvordan er det for en psykolog å anbefale en behandling uten å fullt ut forstå begrunnelsen – og hvordan oppleves dette for pasienten? I ytterste konsekvens kan ansvaret for vurderingen oppleves forskjøvet til en matematisk modell. Bandura (1999) beskriver hvordan ansvarsforskyvning kan skape distanse til konsekvensene av egne handlinger. Dersom klinikerne ikke lenger opplever seg som en aktiv aktør, kan det redusere klinikers emosjonelle ubehag knyttet til prioriteringer. Bateson (1991) påpeker at psykologisk distanse kan føre til svekket motivasjon og handling. KI kan forsterke et problem den var ment å avhjelpe, gjennom å sløve klinikers prioriteringskompetanse. En viktig løsning er at klinikere i fremtiden bør ha forkunnskap om ansvarlig bruk av KI. Spørsmålet vi ved Modum Bad stiller oss er ikke *om* KI skal innføres i vår kliniske hverdag, men *på hvilken måte*. I beste fall kan maskinlæringen faktisk *lære* klinikerne noe nytt om hvilke faktorer vi overser i våre kliniske vurderinger.

Allianse og tillit



For at pasienten skal våge å dele intime og sårbare temaer, må relasjonen bygge på tillit til at opplysninger håndteres sikkert. Personvern og datavern er ikke bare juridiske anliggender, men fundamentale for den terapeutiske alliansen, som belyst av Nissen-Lie & Stänicke (2025). De siste årene har vi sett en rekke KI-skandaler, og helsevesenet er ikke immunt mot feilslått bruk av prediksjonsverktøy. Obermeyer et al. (2019) dokumenterte at en algoritme for vurdering av behandlingsbehov systematisk undervurderte helsehjelpbehovet hos minoritetsbefolkningen, ettersom tidligere helsekostnader ble brukt som indikator. Minoritetsbefolkningen hadde lavere helsekostnader og ble derfor vurdert å ha mindre behov for oppfølging. Dermed fremsto maskinen som nøytral, mens den i realiteten videreførte strukturell ulikhet. Slike eksempler illustrerer faren for at algoritmer kan forvalte og forsterke skjevheter forkledt som objektivitet (Narayanan & Kapoor, 2024). For å redusere risikoen for algoritmisk skjevhet må valg av variabler og evaluering av modeller gjøres i tett samarbeid med de kliniske ekspertene. Syntetiske data – kunstig genererte, men representative datasett – kan være ett virkemiddel for å korrigere disse skjevhetene.

Eierskap til kliniske vurderinger

KI kan bidra til mer effektiv ressursbruk og mer treffsikre behandlingstilbud. Prioriteringsalgoritmer kan redusere feil som oppstår når klinikere uunngåelig preges av menneskelige bias. Samtidig innebærer bruk av KI alvorlige utfordringer knyttet til systematiske skjevheter – som kan være vanskelig å oppdage – og til svekket opplevelse av ansvar. Som Bandura (1999) påpeker, kan moralsk frakobling redusere empati og eierskap til kliniske beslutninger.

Utover at algoritmer bør være transparente og forklare sine vurderinger og formidle den statistiske usikkerheten som følger med, bør klinikeren ha kjennskap til teknologien som brukes og hele tiden insistere på hvor ansvaret ligger. Kunstig intelligens kan fremstå som en fristende snarvei i et presset helsevesen, men feilbarlighet gjelder også maskiner. Tiden vi sparer med KI-revolusjonen, bør på den andre siden bringe refleksjon over hva den åpner opp for skal den styrke menneskelige verdier i terapirommet. Med KI på bakrommet kan vi samtidig forvente at større kapasitet til pasienten, med økt rom for innlevelse, genuin kontakt og profesjonelt skjønn, blir en hoveddel av arbeidshverdagen.

Ved Modum Bad er vi opptatt av å fremheve det etiske i å styrke de evidensbaserte virkemidlene som kan gjøre terapi mer virksom for pasientene.

På den ene siden kan algoritmer spare samfunnet og pasienten fra overbehandling og forhindre behandlingsforløp uten varig effekt. Samtidig kan også prediksjonsalgoritmene komme med anmodninger om større behandlingsintensitet og lengre forløp. Hvilke etiske implikasjoner oppstår hvis det er sprik mellom den økonomiske styringen og prediksjonen fra et framtidig sofistikert, datadrevet og reliabelt KI-system. Hvordan vil vi som fagprofesjon møte stadig mer informerte vurderinger i lyset av begrensede ressurser?

Som forskere på KI i en klinisk setting er vi opptatt av å formidle at fremtiden for KI i slike sammenhenger ikke handler om å erstatte klinikerens skjønn og varme med en prosessor. Det handler mer om å ta i bruk nye verktøy for å styrke det terapeutiske håndverket.

Etisk forsvarlig bruk av KI krever klare ansvarslinjer med mennesket i førersetet, åpne og etterprøvbare modeller og kontinuerlig kvalitetssikring. Med målrettet opplæring av klinikere tenker vi at KI kan bli et supplement som styrker behandlingen uten å undergrave tillit eller faglig skjønn.

Referanser



- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193–209. https://doi.org/10.1207/s15327957pspr0303_3
- Batson, C. D. (1991). *The altruism question: Toward a social-psychological answer*. Hillsdale, NJ: Erlbaum.
- Elvik, I. (2025). Psykologer mangler vurderingsferdigheter. *Tidsskrift for Norsk psykologforening*, 62(1), 44–47. <https://www.psykologtidsskriftet.no/artikkel/2025as01ae-Psykologer-mangler-vurderingsferdigheter>
- Narayanan, A. & Kapoor, S. (2024). *AI snake oil: What artificial intelligence can do, what it can't do, and how to tell the difference*. Princeton University Press.
- Nissen-Lie, H. A. & Stänicke, E. (2025). Fremtidens psykoterapi: Hvilken rolle bør kunstig intelligens spille?. *Tidsskrift for Norsk psykologforening*, 62(9), 512–518. <https://www.psykologtidsskriftet.no/artikkel/2025as07ae-Fremtidens-psykoterapi-Hvilken-rolle-b-r-kunstig-intelligens-spille->
- Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
- Rubel, J.A. & Lutz, W. (Red.) (2025). *Feedback Informed Psychological Therapy: Science and Practice*. Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-90428-8>