Maskinlæring trenger psykologhjelp

Astrid Brænden
psykolog og stipendiat, UiO

Maskinlæringsmodeller som kan predikere atferden vår, utfordrer den psykiske helsen. Psykologer må kjenne sin besøkelsestid.

Astrid Brænden. Foto: privat

Med lanseringen av samtaleroboten ChatGPT kan et kappløp innen maskinlæring, også kalt kunstig intelligens, ha begynt.

En debatt om hvilke konsekvenser teknologi som kan forutsi enkeltindividets atferd (kanskje bedre enn en selv) vil ha for folks psykiske helse, bør vi ha *før*, og ikke etter at teknologiselskaper har utviklet teknologien. Psykologers kunnskap er viktig i debatten om i hvilken kontekst maskinlæring kan være verdifull og forsvarlig – og hvor det ikke vil være det.

Hva er maskinlæring?

Maskinlæring benytter statistiske metoder for å finne mønstre i store datamengder. Tekst, tall, bilder og video kan brukes som data for modellæring.

Et høyaktuelt eksempel er ChatGPT, en maskinlæringsmodell som bygger på språk. Det er en gratis chattetjeneste hvor du kan stille spørsmål og få svar. For eksempel kan du beskrive et psykologisk kasus og be om en behandlingsplan, som i sin form fremstår som å være skrevet av et menneske.

Bruken av ChatGPT har vokst i rekordfart. Etter kun fem dager hadde chattetjenesten mer enn én million brukere. Til sammenligning brukte Netflix og Instagram henholdsvis 3,5 år og 2,5 måneder på å nå det samme tallet.

Psykologers kunnskap er viktig i debatten om i hvilken kontekst maskinlæring kan være verdifull og forsvarlig

Utviklerne av ChatGPT, OpenAI, hevder at chattetjenesten er tilgjengeliggjort uten annet formål enn å få tilbakemeldinger for å lære om dens styrker og svakheter. Dette kan godt være tilfellet, men samtidig betyr det at millioner av mennesker samspiller med tjenesten uten at selskapets intensjoner er tydelige og transparente. Vår kollektive bruk gir tjenesten massive mengder ny data om hva vi er opptatt av, og dermed innsikt i våre tanker.

Dette får meg til å undre: Hvor mange og lange opptak av hva du sier, vil en språkmodell trenge for å forutsi det neste du vil si? Ville en, to eller tre måneder med kontinuerlige opptak av dine daglige interaksjoner vært nok? For de fleste er det sjelden mobilen, som kan gjøre lydopptak, er langt unna. Lyd kan maskinelt og automatisk gjøres om til tekst, og slik bli data for maskinlæring. Mobil og smartklokker kan videre registrere hva du søker etter, dine «klikk», helsedata, observasjonstid, og hva du eksponeres for. Hvis utviklere vil og dette ikke er tilstrekkelig regulert, eller hvis noen vil misbruke teknologien og den mangler tilstrekkelig sikkerhet, er det nærmest ubegrenset hva maskinlæringsteknologi kan lære om deg og bruke til å forutsi atferden din.

Flere hevder at ChatGPT har kickstartet et kappløp innen kunstig intelligens, og Google lanserte 6. februar en tilsvarende språkmodell kalt Bard. Et slikt kappløp er imidlertid ansett som farlig nettopp fordi det kan ramme etiske hensyn, personvern og informasjonssikkerhet.

Konsekvenser for psykisk helse

Hvilke konsekvenser denne typen maskinlæring kan få for psykisk helse, er vanskelig å forutsi. Likevel vil jeg fremheve to mulige konsekvenser, og håper det engasjerer andre til å tenke gjennom flere perspektiver.

For det første: Stadig mer avanserte utgaver og bruksområder av ChatGPT-teknologi kan komme til å utfordre menneskers opplevelse av verdi. Foreløpig er ikke denne teknologien integrert i biologisk materie og har ikke tilgang til emosjoner eller sensoriske opplevelser. Teknologien kan ikke anses for å ha generell menneskelig intelligens eller på noen måte være «mennesker». Opplevelsen av å være nyttig, som er knyttet til opplevelsen av verdi, kan likevel utfordres når algoritmer i økende grad griper inn i hverdagen din, viser seg overlegen i å besvare komplekse spørsmål, eller gjør jobben din. Kanskje kan dette føre til at flere mennesker oppsøker terapi, særlig de som må omstille seg.

For det andre: Modeller som med høy presisjon kan predikere individers atferd, kan utfordre menneskers opplevelse av fri vilje.

Dersom maskinlæring er verktøyet som i størst grad vil kunne hjelpe oss med å predikere atferd og mentale prosesser, må psykologer ha bedre forståelse av slik teknologi

Fri vilje er noe psykologien og filosofien, som psykologien opprinnelig er bygget på, har kunnskap om. Selv om mange lever godt med en forståelse av at fri vilje ikke finnes, lever mange andre

med opplevelsen av at fri vilje er sentralt for deres liv og gir livet mening. I takt med at teknologien gradvis viser hvordan individers atferd og livsløp kan predikeres med høy presisjon, bør psykologer formidle kunnskap om hvordan mennesker kan leve meningsfulle liv uavhengig av hvorvidt fri vilje eksisterer. Det er uheldig hvis mennesker blir deprimerte eller unngår å ta ansvar for livene sine fordi enorme prediksjonsmodeller skaper en forståelse av at de ikke har selvagens. Jeg er altså bekymret for den psykiske helsen til mennesker som kan komme til å oppleve at deres eksistensielle grunnlag, som fri vilje, brått blir revet vekk når de innser at store deler av sine liv kan forutsies.

Kompleksitet og uforutsigbarhet skaper mening og verdi. Dersom opplevelsen av dette blir betydelig utfordret av maskinlæring, kan det ramme menneskers psykiske helse.

Hva vi kan gjøre

Etter mitt syn har maskinlæring i for liten grad blitt drøftet i vårt fagfelt. Psykologi er per definisjon studiet av atferd og mentale prosesser. Dersom maskinlæring er verktøyet som i størst grad vil kunne hjelpe oss med å predikere atferd og mentale prosesser, må psykologer ha bedre forståelse av slik teknologi. Hadde dette blitt tatt på alvor tidligere, ville psykologer kanskje vært i bedre posisjon til å påvirke teknologiutviklingen som individet virker å ha for lite kontroll over i dag.

Psykologer kan fortsatt få en viktig rolle i å avgjøre hvordan maskinlæring bør og ikke bør brukes. Vi har kunnskap om kompleksiteten i å anslå ny atferd basert på tidligere atferd, og kjenner til faktorer som kan påvirke dette og (derfor) bør tas hensyn til. Ettersom prediksjonsmodeller heller ikke forstår hva de selv gjør, er psykologers kunnskap viktig for å kunne si noe om rimelige og urimelige prediksjoner. Vårt fag beskjeftiger seg også med å skape endring og variasjon i etablerte mønstre av språk, tanker og atferd som har blitt uheldige. Psykologer har altså balansert kunnskap om at slike mønstre kan være vanskelig å endre, men at dette er mulig og om hvordan. Denne innsikten er viktig for å vurdere i hvilke kontekster maskinlæring kan være verdifullt og forsvarlig – og hvor det ikke vil være det.

Maskinlæring i terapi

Brukt på en sømmelig måte kan maskinlæring gi behandlere en dypere forståelse av en pasients vansker og bidra til mer persontilpasset behandling. Dette er mulig gjennom maskinlæringsmodeller, som fra store datamengder benytter oppdatert kunnskap om pasientens symptombilde til å «forstå» lidelsen og foreslå behandling. For personer med milde psykiske vansker, kan hjelp bli rimeligere og mer tilgjengelig med virtuelle terapeuter og chattetjenester.

Maskinlæring som et verktøy i terapi, må likevel alltid veies opp mot mulig risiko. Sensitive data kan komme på avveie eller misbrukes. Psykologer bør heller ikke ha ubegrenset tillit til analyser fra maskinlæringsmodeller. For å kunne foreta gode vurderinger må psykologer forstå teknologiens begrensninger. For eksempel kan slike modeller gjøre systematiske feil basert på sitt datagrunnlag. Andre risikoer er hvis behandlere blir mer opptatt av teknologien enn mennesket foran seg. Et eksempel er den amerikanske digitale helsetjenesten for emosjonell støtte kalt Koko. Koko brukte maskinlæring i digital terapi, hvor det var tvilsomt om brukerne visste dette. Selskapet er derfor anklaget for å ha brutt kravet om informert samtykke.

Selv om noen former for maskinlæring kan skape verdi for mennesker, kan utstrakt og uregulert bruk av maskinlæring komme til å ramme menneskers psykiske helse. Maskinlæring brukt til å predikere enkeltindividers atferd med høy presisjon, kan snart bli eller er allerede mulig. Før dette

utvikles eller misbrukes, må psykologer med sin kunnskap om kompleksiteten i atferd og mentale prosesser, ta en sentral rolle i å avgjøre hvor og hvordan maskinlæring kan og ikke kan brukes forsvarlig.

Publisert: 08.02.2023, Tidsskrift for Norsk psykologforening.