

UDVISER SOCIALE CHATBOTS EMPATI?

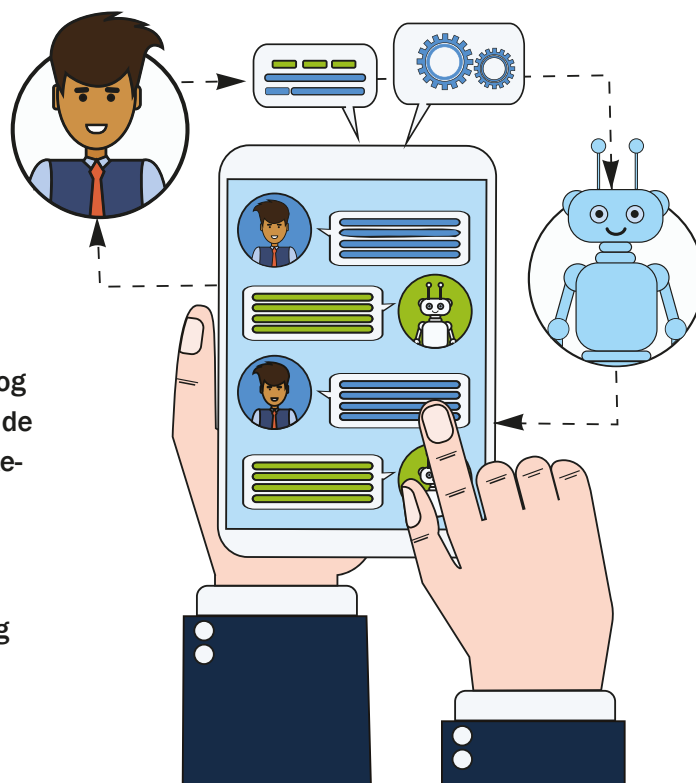


Illustration: Colourbox

Sociale chatbots bliver mere og mere avancerede. Men bliver de også mere og mere menneskelignende i den forstand, at de kan udvise empati?

Umiddelbart er svaret nej, viser en ny specialeafhandling fra Roskilde Universitet.

Selvom alle i dag taler om kunstig intelligens (KI), er det de færreste, der kender omfanget og konsekvenserne af den hurtigtvoksende teknologi. Fænomenet er i rivende udvikling, og hver dag kommer vi et skridt nærmere, hvordan man ved hjælp af nye matematiske beregninger, metoder og teknologier kan efterligne den menneskelige intelligens og hjernens tankeprocesser i en sådan grad, at systemer kan spejle menneskets sociale evner og interagere med os på en menneskelig overbevisende måde. I særlige tilfælde i en sådan grad, at det er svært at skelne mellem,

hvornår vi kommunikerer med et menneske og en maskine. Et eksempel på disse KI-systemer er chatbots, som allerede er en integreret del af mange menneskers hverdag.

Menneskelige maskiner

Chatbots er kunstigt intelligente robotter, som man kan chatte med. De er designet som digitale assistenter, der kan håndtere og hjælpe med for eksempel problemer omkring køb af et produkt på en hjemmeside. De første chatbots så dagens lys i 1960'erne. Siden da, især i løbet af de seneste 10 år, er de blevet mere og mere alminde-

lige i store firmaer som IKEA. Men de kan også benyttes som en virtuel ven for mennesker, der har brug for dialog og selskab i hverdagen. Derfor opstår der naturligt nogle etiske og moralske samfundsdilemmaer omkring mødet mellem menneske og maskine. For hvad betyder det for interaktionen, når det ikke længere er et virkeligt menneske, vi kommunikerer med, men i stedet en maskine, der forsøger at efterligne menneskelig intelligens?

Hvis en chatbot opfører sig menneskelignende i interaktionen med os, er det afgørende, at den

Om forfattere:



Jamilla Lyng, cand. scient. i Informatik og Performance Design, Roskilde Universitet
Forfatter til specialeafhandlingen "Udviser sociale chatbots Theory of Mind?" 2019
jamilla_lyng@hotmail.com



Kathrine Glargaard Andersen, cand.scient. i Informatik og Kommunikation, Roskilde Universitet
Forfatter til specialeafhandlingen "Udviser sociale chatbots Theory of Mind?" 2019
katglaand@gmail.com



Torben Brauner er dr. scient. og lektor i datalogi på Roskilde Universitet. Forsker i logik og anvendelsen af logik i datalogi samt kognitionsvidenskab og filosofi. Han var vejleder på ovennævnte specialeafhandling.
torben@ruc.dk

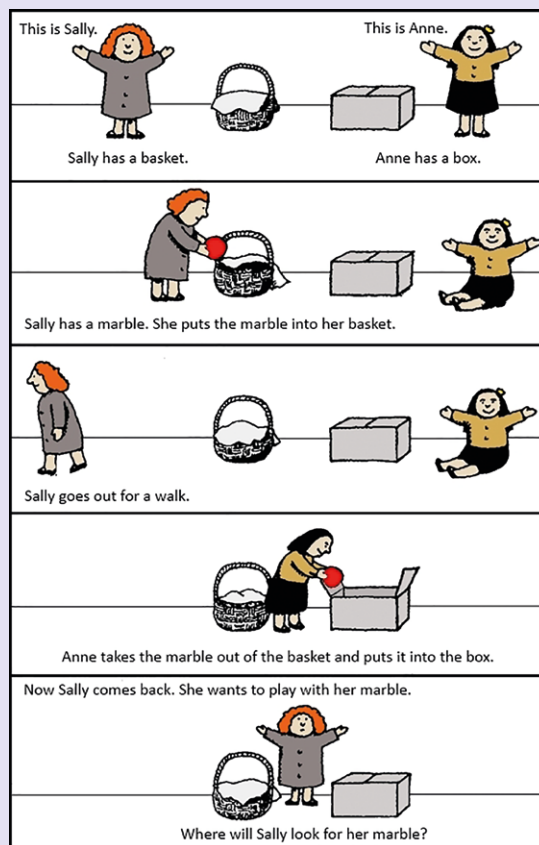
Empati og Theory of Mind

Empati er den menneskelige, sociale evne til at kunne sætte sig i en anden persons sted og opfattelse af virkeligheden og gør mennesket i stand til at mærke og forstå en andens følelse og emotionelle tilstand på egen krop og sind. Det psykologiske begreb Theory of Mind (ToM) blev udviklet af forskerne David Premack og Guy Woodruff i 1978 og er beslægtet med empati. ToM er en socialkognitiv evne til at kunne sætte sig i en anden persons sted og tilskrive både sig selv og andre mentale tilstande såsom overbevisninger ("beliefs"), hensigter, ønsker, viden, etc.

Forskellige variationer af psykologiske tests bruges verden over til at undersøge børns ToM. Psykologiske studier viser samstemmende, at typisk udviklede børn kan bestå disse tests fra omkring fire års-alderen, hvor evnen udvikles.

Den mest anvendte ToM-test er Sally-Anne Testen, som er udviklet af Simon Baron-Cohen, Alan M. Leslie og Uta Frith i 1995. Testen tager udgangspunkt i to dukker, Sally og Anne, som præsenteres for barnet, der testes. Sally placerer en kugle i en kurv og forlader rummet. Anne, som nu er alene, tager kuglen og gemmer den i en kasse. Sally kommer tilbage og barnet bliver spurgt: "Hvor vil Sally kigge efter sin kugle?" For at bestå dette 'Belief Question', skal barnet svare "kurven". Herved giver barnet udtryk for, at Sally har en falsk overbevisning (false belief), da hun ikke ved, at kuglen i virkeligheden er blevet gemt et andet sted. Herefter stilles to kontrolspørgsmål: "Hvor er kuglen i virkeligheden?" (Reality Question) og "Hvor var kuglen til at starte med?" (Memory Question).

Kontrol-spørgsmålene er afgørende for at sikre, at barnet både har viden omkring kuglens reelle, nuværende placering og en nøjagtig hukommelse og forståelse af dens tidligere placering.



er i stand til at udvise en adfærd, der afspejler nogle af menneskets socialkognitive egenskaber såsom emotionel genkendelse, empati, moral, viden om sociale normer etc. Hvis ikke chatbotten er designet til at forholde sig til disse emotionelle faktorer i dialoger omkring eksempelvis psykiske problemstillinger, er der en risiko for, at de kan vildlede mennesker og rådgive dem forkert i en sårbar og psykisk ustabil situation, hvor mennesket er særlig modtagelig overfor chatbottens eksakte råd og formuleringer. Hvis et menneske tager imod psykologisk behandling eller rådgivning fra et individ, som ikke forstår eller kan sætte sig i menneskets psykiske tilstand eller sted, må det forventes at kunne medføre betydelige konsekvenser

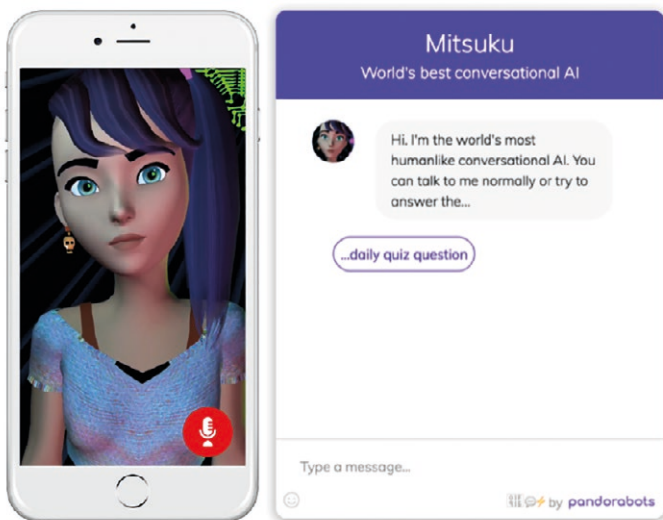
og gå ud over menneskets psykiske ve og vel.

I netop denne sammenhæng er en evne som empati, der er nødvendig for at kunne forstå samt sætte sig ind i andres følelser og sted, særlig vigtig for chatbots. Men hvordan kan man undersøge, hvor menneskelignende en chatbot i praksis opfører sig i interaktionen med mennesker? Det kan man ved hjælp af den såkaldte Turing-test, opkaldt efter matematikeren og logikeren Alan Turing (1912-1954), der i øvrigt er den moderne datalogis fader.

Turing-testen er en metode til at afgøre, om en maskine udviser intelligens som et menneske (Turing-testen afgør således ikke, om

en maskine virkelig er intelligent, for eksempel om den har bevidsthed). Turing-testen går ud på, at et menneske og en maskine er placeret i hvert sit lukkede rum, og en dommer (et menneske) uden for rummene skal forsøge at afgøre, i hvilket rum henholdsvis mennesket og maskinen befinder sig. Dommeren må spørge om hvad som helst, men har kun elektronisk kommunikation til rådighed, der ikke afslører, om man kommunikerer med en maskine eller et menneske.

Turing-testen har siden 1990 været anvendt i den såkaldte Loebner Prize konkurrence, der er en formel realisering af Turing-testen. Konkurrencen blev lanceret af opfinderen Hugh Loebner med formålet om at finde det mest menneskelignende



Chatbotten Mitsuku har vundet Loebner Prize konkurrencen fem gange.

kunstigt intelligente chatbot-system, som aktuelt er på markedet.

Hvor langt er vi nået i dag?

Chatbotten Mitsuku har vundet Loebner Prize konkurrencen hele fem gange og markedsføres i dag som verdens mest menneskelignende chatbot. Derfor må Mitsuku blandt andet forventes at udvise tegn på empati i en skriftlig dialog med et menneske.

Netop det har to af denne artikels forfattere (Kathrine Glargaard Andersen og Jamilla Lynge) undersøgt i en specialeafhandling i informatik fra Roskilde Universitet. Konkret undersøgte specialet, om chatbotten Mitsuku er i stand til at udvise Theory of Mind (empati) på niveau med typisk udviklede børn i fire års-alderen, da det typisk er omkring denne alder, at mennesker udvikler empatiske egenskaber.

Undersøgelsen af Mitsuku blev udført med udgangspunkt i Turing-testen, som den anvendes i Loebner Prize konkurrencen, hvor man gennem skriftlig kommunikation med chatbotten søger nogle bestemte svar. Det er derfor en brugbar testmetode i forhold til at spore nogle empatiske træk i interaktionen mellem et menneske og et chatbot-system.

I specialet blev problemstillingen grebet empirisk an ved at udsætte

Mitsuku for en række standard psykologiske Theory of Mind-tests, der bestod af skriftlige fortællinger med tilhørende spørgsmål på fem forskellige niveauer med stigende sværhedsgrad.

På trods af, at testene på forskellig måde blev indskrænket og tilpasset chatbottens teknologiske muligheder og kunnen, lykkedes det kun i meget begrænset omfang for Mitsuku at udvise en grad af Theory of Mind, der udvises af typisk udviklede fire-årige børn. Det skyldes formentlig to ting.

Chatbotten havde for det første svært ved at forstå fortællinger med flere sætninger, uanset hvad de handlede om. Når en sætning indeholder mange ord og betydninger, er Mitsuku kun i stand til at håndtere forståelsen af godt én sætning ad gangen. Det gjorde testene, som

tog udgangspunkt i små fortællinger, særligt udfordrende for Mitsuku at formulere korrekte svar til – men også for Kathrine og Jamilla, da testspørgsmålene skulle reduceres til en længde, hvor Mitsuku kunne holde styr på det, hvilket ofte var vanskeligt. For det andet har Mitsuku en tendens til at “efterabe”, forstået sådan, at systemet kopierer tekststykker, man allerede har skrevet til den i et testspørgsmål. Man kan altså ikke vide sig sikker på, om Mitsuku selv formulerer sine svar ud fra en reel forståelse, som pudsigt nok ligger utrolig tæt op ad formuleringer, man allerede har “fodret” systemet med tidligere – eller om chatbotten rent faktisk kopierer de ord og input, den tidligere har fået. Det skabte store fejlkilder, da man ikke kan være sikker på, om chatbotten reelt udviser empati, eller om den blot taler en efter munden. Selvom specialets studier tog højde for disse faldgruber, for eksempel ved oftest at anvende korte, skriftlige fortællinger, viste der sig kun ringe sammenhæng mellem Mitsukus status som verdens mest menneskelignende chatbot og det, at kunne udvise sociale evner som empati.

Nok er kunstig intelligens gennem årene blevet dygtigere til at kommunikere med os og forstå vores sprog, men der synes endnu ikke at være udviklet chatbots med en generel sproglig kapacitet, så chatbots kan udvise menneskelige socialkognitive evner som Theory of Mind – heller ikke selvom de markedsføres som “menneskelignende”. ■

Loebner Prize konkurrencen

- Konkurrencen blev lanceret i 1990 og afholdes én gang årligt.
- Selve konkurrencen består af fire runder, hvor hver chatbot parres med et menneske. Fire dommere chatter med hvert par (en chatbot og et menneske), hvorudfra de skal afgøre, hvem der er chatbotten, og hvem der er mennesket.
- Til sidst rangerer dommerne de fire chatbots efter, hvor menneskelignende de formår at svare sammenlignet med menneskets svar.
- Førstepladsen tildeles chatbot-systemet med den højeste samlede score.

Mere information
Andersen, K. G. & Lynge, J. (2019). *Udviser social chatbots Theory of Mind?* Specialeafhandling i Informatik, Roskilde Universitet. Tilgængelig via Det Kongelige Bibliotek.

Turing, A. M. (1950). *Computing Machinery and Intelligence* i *Mind*. Oxford University Press: Mind Association. S. 433-460

Worswick, S. (2018b). *Mitsuku wins Loebner Prize 2018!* <https://medium.com/pandorabots-blog/mitsuku-wins-loebner-prize-2018-3e8d98c5f2a7>

Wellman, H. M. and Liu, D. (2004), *Scaling of Theory-of-Mind Tasks*. *Child Development*, 75: 523-541.