

# Forord

2023 var året hvor kunstig intelligens (AI) og *large language models* (LLM) for alvor satte deres aftryk på almindelige sprogbrugeres liv og i særdeleshed på forskning og undervisning. Chatbotten ChatGPT genererede avisoverskrifter (i begge betydninger) og truede med (afhængig af standpunkt) at revolutionere eller undergrave forskning, undervisning og eksamination. Pludselig blev vi opmærksomme på at det at skrive en sammenhængende og forståelig tekst ikke krævede et menneske endsiges en lang uddannelse.

Nu, i 2024, er vi så småt ved at vænne os til den nye verdensorden, og vi bevæger os fra spekulation til levede erfaringer. LLM'er er ikke længere en hypotetisk *disruption* af undervisning eller arbejdsmarked. LLM'er har fundet vej ind i undervisningslokaler fra folkeskole til universitet. Vi diskuterer nu ikke længere hvordan vi skal dæmme op for snyd med hjælp fra AI, men hvordan studerende skal undervises i at bruge LLM'er i lærings- og eksamenssammenhæng. Vi taler mindre om hvordan LLM'er vil overtage humanisters arbejdsfunktioner, og mere om hvordan vi træner humanister til en fremtid hvor de skal arbejde tæt sammen med AI.

Vi begynder måske også at få øjnene op for de mere interessante og foruroligende spørgsmål som AI og LLM'er kaster af sig. Hvad siger LLM'er fx om menneskets sprogevne og menneskets status som den eneste art der bruger sprog? Med GPT4 er vi vidne til at sprogrelaterede opgaver nu bliver løst på en måde der ofte ikke længere kan skelnes fra hvad der er produceret af mennesker. Vi begynder at nærme os drømmen om *Artificial General Intelligence* – maskinen der kan løse en lang række kognitive opgaver på niveau med et menneske. Hvis maskinen løser opgaver tilsyneladende som et menneske, er det så forkert at sige at maskinen 'tænker'? Eller omvendt: Hvis maskinen ikke tænker, hvorfor siger vi så at mennesker 'tænker'? Vi vil ikke gå for langt ind i denne filosofiske debat, som vækker mindelser om tidligere tiders debatter om behaviorisme vs. kognitivismen. Men vi peger på den fordi den er rele-

vant for terminologien i bidragene til dette nummer: Nogle bidrag bruger en terminologi som italesætter LLM'er som 'en maskine'. De taler fx om "tekster der er genereret" eller om "maskinens output". Andre bidrag bruger en terminologi som italesætter LLM'er som en intelligent aktør. De taler fx om at LLM'er "ræsonnerer" eller "forstår". I redaktionen finder vi denne variation fascinerende og illustrativ for det tidspunkt vi befinder os på i vores forståelse af LLM'er.

Under overskriften *Sproget, mennesket og teknologien* kan vi denne gang bringe fire artikler og to anmeldelser. De tre første artikler går meget direkte til spørgsmålet om LLM'er og deres evne til at producere sprog som ikke afviger fra det et menneske kan.

Bolette Sandford Pedersen, Nathalie C. Hau Sørensen, Sussi Olsen og Sanni Nimb giver i deres artikel *Evaluering af sprogforståelsen i danske sprogmodeller – med udgangspunkt i semantiske ordbøger* forslag til hvordan man kan afprøve og evaluere, *benchmarke*, danske LLM'ers evne til at håndtere tekstsemantisk-logiske opgaver som bl.a. synonymhåndtering, inferenser og følgeslutninger, eller, med forfatterens egne ord, LLM'ernes evne til at 'ræsonnere' og 'forstå'. Evalueringerne er baseret på en række eksisterende danske ordbøger og teknologiske sprogresurser som flere af forfatterne har udviklet over de seneste år. Forfatterne konkluderer at deres benchmarksystem er tilstrækkeligt udfordrende for de testede LLM'er, at de 'begår fejl', men også at ChatGPT 4.0 præsterer rigtigt godt.

I artiklen *Kan AI reproducere fagdisciplinær stemme? Et komparativt korpusbaseret studie af GPT4's evne til at reproducere fagdisciplinær stemme i AI-genereret sprogvidenskabelig prosa* sammenligner Ea Lindhardt Overgaard og Ulf Dalvad Berthelsen et korpus af sprogvidenskabelige tekster skrevet af GPT 4.0 med et korpus af menneskeskrevne sprogvidenskabelige tekster. De to korpusser sammenlignes for deres 'stemmes' *stillingtagen*, *engagement* og brug af *fagspecifikt ordforråd*. Analysen viser at GPT i høj grad er i stand til at efterligne tekster skrevet af mennesker, eller sagt på en anden måde, at den har lært at udtrykke sig med en passende akademisk stemme.

Jonas Nygaard Blom, Alexandra Holsting og Jesper Tinggaard Svendsen undersøger i deres artikel *På sporet af chatbottens sproglige fin-*

geraftryk. En sproglig ophavsanalyse af tekster skrevet af danskstuderende og ChatGPT om det er muligt at skelne sprogligt mellem tekster skrevet af nystartede danskstuderende på universitetet og ChatGPT. Forfatterne sammenligner blandt andet de to tekstkorpussers forskellige mængde af afvigelser fra retskrivningsnormen, udbredelsen af interaktionelle resurser samt afsnits- og periodelængde. Forfatterne konkluderer at hvis de studerende og ChatGPT får stillet samme opgave, har chatbotten dels en mere ensartet måde at skrive på, dels færre ortografiske fejl end de studerende. Det er dog ifølge forfatternes undersøgelse muligt i en vis udstrækning at sløre at en tekst er skrevet af ChatGPT ved at prompte den til at efterligne de studerendes skrivestil.

Tina Thode Hougaards bidrag *Emojier som parasproglige resurser i skrevet onlineinteraktion* skiller sig ud fra de øvrige ved ikke at handle om GPT og LLM, men i stedet om menneskers brug af teknologiens muligheder til at opfylde deres kommunikative behov. Computermediert kommunikation var i udgangspunktet tekstbåren, men med tiden udviklede sprogbrugere forskellige former for ikoniske tegn som ”:-)” som senere blev konventionaliseret med et stigende antal *emojier*. Hougaard undersøger det kommunikative potentiale for den slags tegn og konkluderer at de samlet set tjener den funktion at få den talendes krop *virtualiseret* i den computermedierede samtale.

De to sidste bidrag til dette nummer af NyS er Ib Ulbæks, lektor emeritus, anmeldelse af Peter Widells bog *Sproget i verden. Tekstforståelsens filosofiske grundlag* (udgivet i 2023) og Holger Juuls, leder af Center for Læseforskningen ved Københavns Universitet, anmeldelse af Christian Becker-Christensens bog *Dansk Retsskrivning: Bogstav – Lyd – Bogstav. En grafemisk beskrivelse af dansk ortografi* (udgivet i 2021). Anmeldelser der optræder i temanumre af NyS, er ikke i udgangspunktet knyttet til temaet.

Vi ønsker læserne god læselyst!