FOLKETINGET

Folketingstidende
Tillæg A

Beslutningsforslag nr. **B 101** Folketinget 2023-24

Fremsat den 19. december 2023 af Lisbeth Bech-Nielsen (SF) og Karina Lorentzen Dehnhardt (SF)

# Forslag til folketingsbeslutning

## om at skabe en stor dansk sprogmodel, også kaldet LLM eller large language model

Folketinget pålægger regeringen i indeværende folketingssamling at igangsætte arbejdet med at udvikle en dansk generativ sprogmodel, bl.a. med brug af data, der er tilgængelige hos Det Kgl. Bibliotek og andre større vidensdatabaser.

# *Bemærkninger til forslaget*

Forslagsstillerne ønsker, at Danmark udvikler en stor dansk sprogmodel, også kaldet Large Language Model, LLM, som vi kender fra bl.a. OpenAI's ChatGPT.

Både i EU-regi og nationalt i Danmark har vi nogle særlige værdier og regelsæt, som de amerikanske sprogmodeller ikke tager højde for. Derfor mener forslagsstillerne, at vi nationalt bør arbejde for at være uafhængige af udenlandske kommercielle interesser og have fokus på datasikkerhed, transparens og dansk indhold i en dansk sprogmodel.

Andre lande, bl.a. vores skandinaviske naboer, er langt foran os og har fået øjnene op for potentialet i kunstig intelligens, og i Sverige er man gået i gang med at udvikle en sprogmodel på svensk (»Sverige er i færd med at udvikle sin egen AI sprogmodel«, Dansk IT Sikkerhed, den 11. oktober 2023, og »Skal vi stole på de amerikanske chatbots eller øh ... bare lave vores egne? «, www.zetland.dk, den 9. oktober 2023).

Hvis vi vil sikre os, at kunstig intelligens kan anvendes til fulde i en dansk kontekst uden risiko for større fejl, bør vi udarbejde en dansk sprogmodel, som kan tage højde for danske kulturelle normer, værdier, love og regler. Behovet for at udarbejde en stor dansk sprogmodel er presserende, og det bør derfor sættes i værk snarest muligt. Det handler basalt set om at redde det danske sprog og den danske kultur.

Vi kender i forvejen OpenAI's ChatGPT, der bliver mere og mere udbredt og anvendes i et fortsat bredere perspektiv. ChatGPT kan meget, men den amerikanske sprogmodel tager ikke højde for danske kulturelle normer og værdier, der ikke indlejres i amerikanske sprogmodeller. Hvis man f.eks. spørger chatbotten, om man må stille en barnevogn foran en café, vil den på det kraftigste fraråde dig det og pointere, at du desuden kan blive straffet for det. Den kender ikke til de uskrevne danske regler eller dansk lovgivning.

Der er forskel på, om en sprogmodel primært kender til thanksgivingtraditionen, amerikansk fodbold, 4. juli-fejring og pickuptrucks, eller om dens primære udgangspunkt er juleaften, håndbold, den 5. juni og christianiacykler. Det ene vidensgrundlag er ikke nødvendigvis bedre end det andet, men det skaber to forskellige udgangspunkter og dermed en forskel på det produkt, vi får, når vi anvender sprogmodellen.

Forslagsstillerne mener, at vi bør se generativ AI og tilsvarende fremtidige teknologier som kritisk infrastruktur på lige fod med f.eks. veje, vand og elforsyning. Derfor bør en dansk sprogmodel på en eller anden måde være ejet af fællesskabet. Forslagsstillerne ser helst et offentligt ejerskab eller et offentlig-privat samarbejde.

Forslagsstillerne mener ikke, at det vil være sikkerhedsmæssigt forsvarligt at overlade vigtig ny digital infrastruktur til udenlandske techgiganter og fortsætte med udelukkende at anvende en sprogmodel, som vi ikke har nogen kontrol over.

Sprogmodeller er noget relativt nyt, men de og andre kunstig intelligens-produkter er helt sikkert kommet for at blive. Teknologien er hverken ond eller god, men vi er nødt til at sikre demokratisk kontrol med teknologien.

Bl.a. står vores uddannelsesinstitutioner over for at skulle forholde sig til kunstig intelligens og brugen af det i undervisning, forskning og andet. Det understreger også behovet for en danskbaseret national sprogmodel, hvor datasikkerhed og indhold er i trygge hænder.

*Træning af en sprogmodel*

En sprogmodel skal fodres med enorme mængder tekst. Her foreslår forslagsstillerne at lave et samarbejde med Det Kgl. Bibliotek, lex.dk og andre, der må ses som en ideel partner i forhold til træningsdata i form af tekst og videnskabelige artikler. Brug af træningsdata skal selvfølgelig være under forudsætning af, at de rettighedsmæssige forhold kan afklares. En dansk sprogmodel skal desuden være præget af transparens, så vi i det omfang, det er muligt, kan forstå, hvilket datagrundlag der ligger bag. Det samme gør sig gældende for den algoritme, der skal danne grundlag for en dansk sprogmodel.

*Økonomi og finansiering*

Det anslås, at udviklingen og etableringen af en open source-baseret dansk basismodel vil koste omkring 40 millioner danske kroner. Beløbet er anslået ud fra, at det har kostet OpenAI i omegnen af 5 mio. dollars at udvikle ChatGPT3 i 2020 (»How much did GPT-3 cost?«, www.pcguide.com, den 11. august 2023).

Omkostningerne til udvikling af en dansk sprogmodel kan variere alt efter forskellige modeller. Forslagsstillerne er villige til at drøfte forskellige udformninger og dermed også omkostninger og finansiering.

En stor del af omkostningerne vil være til computerkraft, da det kræver store mængder data at træne og vedligeholde en stor sprogmodel. Det vil være muligt at leje denne kapacitet på europæiske servere, hvilket vil give en høj grad af fleksibilitet og plads til nye træningssæt. En stor dansk sprogmodel vil også kunne etableres på nationalt placerede servere, men med mindre fleksibilitet som omkostning.

Forslagsstillere ønsker at finansiere en dansk sprogmodel gennem det økonomiske råderum, men er åbne for at diskutere andre finansieringsformer. SF har i sit finanslovsudspil peget på en række forskellige finansieringskilder, som forslagsstillerne er villige til at drøfte.

## *Skriftlig fremsættelse*

**Lisbeth Bech-Nielsen** (SF):

Som ordfører for forslagsstillerne tillader jeg mig herved at fremsætte:

*Forslag til Folketingsbeslutning om at skabe en stor dansk sprogmodel, også kaldet LLM eller large language model*

(Beslutningsforslag nr. B 101)

Jeg henviser i øvrigt til de bemærkninger, der ledsager forslaget, og anbefaler det til Tingets velvillige behandling.