

ChatBib

ChatBib – et sikkert redskab til teknologiforståelse og innovation i bibliotekssektoren

ChatBib er et prototype-projekt der viser hvordan generativ AI kan bruges i bibliotekssektoren. Det er inspireret af ChatGPT og andre lignende produkter, der udstiller en chat-grænseflade og giver mulighed for dialog. ChatBib giver mulighed for at afprøve forskellige tilgange til at anvende AI. Der er fokus på kompetenceopbygning, erfaringsdeling og eksperimenter.

Generativ AI fylder meget i medierne og har også givet anledning til interesse og initiativer i bibliotekssektoren, men der mangler et relevant eksperimentarium. ChatBib kan bidrage til dette. Det skal understøtte den dialog, der er fundamentet for at udvikle egentlige AI-produkter i biblioteksregi. KOMBIT og DBC ønsker at fremme anvendelsen af AI i bibliotekerne gennem tekniske løsninger og konkret vidensformidling om AI.

Tilgængeligheden af åbne modeller gør det muligt at tilbyde bibliotekerne et konkret projekt, der genbruger de kompetencer og ressourcer, der allerede findes hos KOMBIT og DBC. Derfor lancerer KOMBIT og DBC nu ChatBib, som gør det muligt for bibliotekerne at blive klogere på og tage skridt til at integrere AI-teknologier i det daglige arbejde. Det sker i et sikkert og transparent miljø uden at data indgår i tech-giganternes forretningsmodeller. Projektets formål er at erfare potentialer og begrænsninger og anvende erfaringerne i udviklingen af fremtidige AI-løsninger i bibliotekssektoren for både borgere og professionelle. ChatBib er et fuldkommen åbent projekt, som biblioteker kan anvende, drage nytte af og påvirke.

Kommercielle AI-chat-løsninger som fx ChatGPT vil stadig være relevante til eksperimenter for biblioteksprofessionelle og til privat brug. De vil også kunne bruges i forskellige former for backoffice. Det offentlige Danmark kan imidlertid ikke bygge på disse løsninger i et borgerperspektiv i forhold til privatliv og transparens.

Håndteringen af personfølsomme oplysninger i biblioteksløsninger skal opretholde de højeste standarder for databeskyttelse. Det er afgørende for at opretholde tilliden til bibliotekerne. ChatBib afvikles i et beskyttet miljø i Danmark og sikres mod enhver form for GDPR-afvigelse samt uønsket deling. I første omgang er det et værktøj, der bruges i et mere eksperimentelt og professionelt rum. Det er et afgørende fokus i ChatBib at udvikle en model, der beskytter privatlivet i en digital tidsalder med ekstrem vækst i kommercielle AI løsninger.

Vi ser frem til dialog, erfaringsudveksling og enhver form for feedback. Dette skal danne grundlag for udvikling af prototyperne på vej mod borgerrettet produktudvikling. Centralbibliotekernes visionsproces for AI-anvendelse og udnyttelse i bibliotekerne forventes ligeledes at kunne påvirke produktudviklingen henad vejen; og tilsvarende forventes ChatBib at kunne give input til visionsarbejdet.

ChatBib udstiller en grænseflade i naturligt sprog ovenpå en stor sprogmodel og fokuserer på biblioteksperspektivet. Projektet vil også kunne anvendes som inspiration til andre offentlige områder og vil kunne indgå i samarbejdsprojekter på tværs af fagsystemer i kommunerne.

ChatBib har fire prototyper indbygget. De viser hver især forskellige muligheder og begrænsninger.

Formålet er på kort sigt at:

- Igangsætte konkrete AI-eksperimenter på bibliotekerne – ved at afprøve prototyperne
- Styrke erfaringsudveksling og kompetenceopbygning i bibliotekerne ved at dele viden om eksperimenterne og afprøvningerne

- Skabe et sikkert økosystem for biblioteksrettede AI-aktiviteter uden de problemer der følger med tech-giganternes forretningsmodeller og datapraksis
- Vise muligheder og begrænsninger i sprogmodeller i praksis – afdække hvor de fungerer godt og hvor de ikke fungerer

Kom igang

Prøv ChatBib: <https://beta.chatbib.dk/>

Under punktet, Kom i gang – er der samlet et gennemgående eksempel fra de 4 prototyper. Dette kan bruges som inspiration.

Send jeres feedback til feedback.chatbib@dbc.dk, eller direkte i ChatBib. Skriv hvad der fungerer og hvad der ikke gør. Giv eksempler og del jeres overvejelser. Vi vil løbende samle feedback og dele opsamlinger og indsigter på ChatBib-hjemmesiden. Her vil surveys også kunne publiceres. Vi forventer at afholde et webinar i løbet af vinteren hvor vi samler op på projektet. Særligt interesserede biblioteker, der ønsker at komme tættere på projektet, kan kontakte Christian Boesgaard (cbo@dbc.dk).

ChatBib

Mere om projektet

For at understøtte udviklingen af AI-produkter i bibliotekerne har KOMBIT og DBC lanceret ChatBib. Projektet er designet til at fremme anvendelsen af AI-teknologier i biblioteker gennem praktiske eksperimenter og vidensdeling. ChatBib udnytter åbne AI-modeller i et sikkert og transparent miljø, som ikke indgår i tech-giganternes forretningsmodeller, hvilket beskytter brugernes data.

Formålet er at understøtte kompetenceudvikling i bibliotekerne og udforske potentialer og begrænsninger ved AI for at kunne udvikle fremtidige AI-løsninger i bibliotekssektoren. ChatBib er åbent for alle og kan bruges til både professionel udvikling og inspiration i andre offentlige sektorer. Projektet sigter også på at sikre høje standarder for databeskyttelse og privatliv.

Der er lavet en chatgrænseflade, der giver adgang til generelle sprogmodeller. Der tages afsæt i design for etablerede produkter som ChatGPT og erfaringer fra DBCs arbejde med SkoleGPT.dk, som er baserede på flere forskellige sprogmodeller i Open Source. Det vil blive videreudviklet til bibliotekssektoren med prompts rettet mod bibliotekarere. Dette er eksemplificeret i den nuværende prototype, men ud fra feedback vil det blive tilpasset.

ChatBib, professionel dialog – ikke borgeranvendelse

Formålet er at igangsætte konkrete AI-eksperimenter sammen med bibliotekerne, styrke den faglige dialog og kritisk vurdere sprogmodellers potentialer og faldgrupper. Senere kan det danne basis for chatflader rettet imod borgere ovenpå biblioteksmaterialer, bibliotekernes viden og fakta, som bibliotekerne kan stå inde for. ChatBib er ikke et offentligt produkt. Prototypens målgruppe er udelukkende bibliotekernes ansatte, ikke borgere. ChatBib er dog ikke hemmelig, så del bare linket med andre.

Vær kritisk

Sprogmodellerne er trænet på gigantiske datamængder. Problemet er, at sprogmodeller og data ikke altid er pålidelige. Sprogmodellen er også kreativ og kan give svar, der er faktisk helt forkerte. Hvis du vil bevise, at modellen fejler, kan du helt sikkert finde masser af svar, som du selv kan besvare bedre gennem åbne kilder på nettet eller på anden vis. Du kan nemt få den til at "lyve". Dette gør dog ikke ChatBib ubrugelig. Der kan nemlig findes masser af relevant og korrekt viden. Den skal tjene sit formål: at være et relevant eksperiment. Derfor opfordrer vi dig til at bruge din kritiske sans, samle erfaringer op og dele med andre. Modellen bliver helt sikkert bedre og bedre henad vejen. I alle tilgængelige sprogmodeller kan man få svar der er uetiske, forkerte og unøjagtige. Du kan hjælpe med henad vejen at fintune og indhegne modellen i ChatBib.

Teknologien

ChatBib anvender Mixtral 8*7B, en sprogmodel med 56 milliarder parametre, som sammen med brugerens kontekst bruges til at generere et svar.

Mixtral og Mistral er sprogmodeller udviklet og udgivet af Mistral AI, et fransk firma, der har som mål at gøre sprogmodeller tilgængelige som åbne værktøjer, der kan bruges til gavn for samfundet uden at være afhængige af de traditionelle forretningsmodeller fra de store teknologiselskaber.

Mixtral 8*7B kører på GPU'er i DBC Digital's datacentre.

ChatBib omfatter også en bred pakke af AI-software, udviklet af DBC Digital, som ejes af KL.

Hvem står bag?

De tekniske partnere i ChatBib er DBC Digital og KOMBIT. En del biblioteksansatte har desuden allerede sagt ja til at deltage i projektet. DBC Digital har mere end 10 års erfaring med udvikling og drift af AI-baserede systemer og machine learning. Vi arbejder med sprogmodeller i relation til metadataskabelse, søgning, recommendere og chatgrænseflader. KOMBIT og DBC Digital er partnere i DSK, Dansk Sprogmodel Konsortium. DBC Digital arbejder sammen med flere universiteter og offentlige virksomheder i relation til AI.

Integration til FBI

Der er arbejdet på at gøre prototypen til en såkaldt RAG (Retrieval-Augmented-Generation), således at der søges i udvalgte dele af FBI-brønden, fx bibliotek.dk. Bibliotek.dk er den ene af de fire tilgange til ChatBib.

Privatliv og GDPR

ChatBib samler ikke data op om dig. Parterne bag ChatBib har et kompromisløst fokus på privatliv og GDPR. Sprogmodeller og Chat afvikles på DBCs datacentre i Danmark.

ChatBib er et system baseret på AI og som bruger skal du derfor vide, at det er en maskine du interagerer med, og den tekst du ser, er genereret af AI.

Fintuning

På nuværende tidspunkt er modellen ikke fintunet og opererer i sin rå form. Dette betyder, at den endnu ikke er tilpasset specifikt til vores behov og stadig er baseret på den generelle træning, den har modtaget. DBC kommer til at træne og fintune ChatBib med relevante datasæt. Vi vil også inkludere brugerfeedback. På et senere tidspunkt etableres en løsning, hvor du som bruger kan sige ja til, at dine spørgsmål, svar og feedback bliver brugt som træningssæt.

Energiforbrug og bæredygtighed

Parterne bag ChatBib er meget optaget af bæredygtighed i forhold til energianvendelse og CO2 aftryk. Et hovedproblem med de store sprogmodeller er at de både ved etablering, tuning og anvendelse kræver ressourcekrævende GPU-hardware. ChatBib vil løbende publicere information om energiforbrug. Alle DBC Digital's datacentre bruger 100% grøn strøm.