BIAS

Bias refererer til forudindtaget eller uretfærdig favorisering eller diskrimination mod visse grupper eller idéer.

Chatbots kan være bias af flere årsager. Her er nogle af de vigtigste årsager til, at chatbots kan have bias.

Menneskelige skabere: Mennesker, der træner chatbots, kan have deres egne forudindtagede holdninger og værdier, som kan påvirke, hvordan chatbot'en reagerer på spørgsmål og anmodninger. Dette kan utilsigtet føre til bias i chatbot'ens adfærd

· Fodring af viden gøres op i mængder, som udgør en større eller mindre sandsynlighed for at indgå i et svar fra en chatbot.

· Farveblanding i metermål kan illustrere overvægt enkelelementer i væsker

· Demokrati: Hvad sker, når er overvægt af vælgere definerer en dagsorden - Lovgivning, økonomi etc.

Træningsdata: Chatbots trænes ofte ved hjælp af store mængder tekstdata fra internettet. Disse data kan indeholde bias, da de afspejler de fordomme og holdninger, der findes i samfundet.

· Hvis dataene indeholder fordomme eller diskrimination, vil chatbot'en lære at gentage disse fordomme. Religiøse og samfundsmæssige problemstillinger, som er værdiladede og drevet af normer og etik er oplagte at arbejde med. Skønlitteraturen på dansk og sprogfagene kan bruges til at sammenholde udsagn fra chatbots gennem analysearbejde.

· Mængden af data kan sammenlignes med pixel i et billede – desto flere pixel jo tydeligere et billede

· Man kan også arbejde med koncepterne: Ekkokamre, censur, ytringsfrihed,

Selektiv træning: Nogle gange trænes chatbots selektivt ved at bruge bestemte typer data, der kan have en skjult bias.

· Hvis en chatbot primært trænes på tekster fra bestemte nyhedskilder eller hjemmesider, kan den udvikle en ensidig opfattelse af emner og tendenser.

· Her vil man kunne trække på bl.a. historie, samfundsfag, kristendomskundskab, hvor man taler ind i de forfattere, som skriver historien og vælger vinkler, religiøse skrifter, som indeholder forkyndende og adfærdsregulerende indhold samt værdiladet og normdefinerende retningslinjer. Hvordan vil disse se ud, hvis de enkeltstående blev lavet om til datasæt?

Sprogbrug og kulturel kontekst: Chatbots kan have svært ved at forstå den komplekse sprogbrug og kulturelle nuancer i sprog. Dette kan føre til fejlagtige eller stærkt forenklede svar, der ikke tager hensyn til den mangfoldighed af udtryk og holdninger i sproget.

· Forskellige udtryksformer indenfor kunsten, herunder sange, spokenword, viser, salmer, ordsprog, dialekter, stand-up, satire og sproglige virkemidler (anafor, korte sætninger, lavt stilleje, sammenligning og ironi), som påvirker modtageren

· Hvis træningsdataen ikke er mangfoldig og repræsenterer forskellige perspektiver og baggrunde, kan chatbot'en have svært ved at give en nuanceret forståelse indenfor et emne eller kulturel norm.

Bias i sprogmodeller kan have en negativ indvirkning på vores samfund. Det kan føre til diskrimination, fordomme og misforståelser. Det er derfor vigtigt at være opmærksom på bias i sprogmodeller og tage skridt til at reducere det. Her er nogle specifikke eksempler på, hvordan bias kan reduceres:

Bias i sprogmodeller kan have en negativ indvirkning på vores samfund…

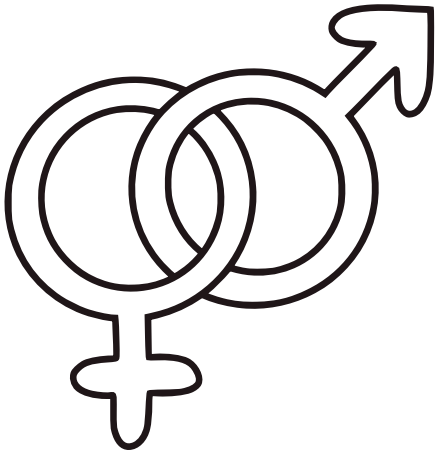
Det kan føre til diskrimination, fordomme og misforståelser. Det er derfor vigtigt at være opmærksom på bias i sprogmodeller og tage skridt til at reducere det. Her er nogle specifikke eksempler på, hvordan bias kan reduceres:

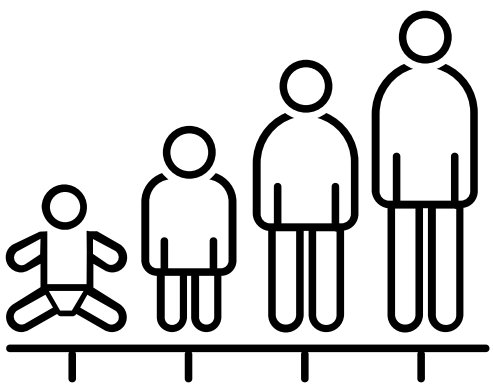
- Et firma, der udvikler en sprogmodel til at generere tekst, kan overveje at bruge en datasæt, der er indsamlet fra en række forskellige kilder. Dette vil hjælpe med at reducere risikoen for, at sprogmodellen bliver skæv i forhold til et bestemt emne.

Det er vigtigt at bemærke, at der ikke findes en enkelt løsning til problemet med bias i sprogmodeller. Det er nødvendigt at tage en række forskellige tiltag for at reducere bias i sprogmodeller og sikre, at de er retfærdige og rimeligt for alle.

Her er nogle eksempler på emner, som man kan arbejde med i undervisningen:

- Køn: Sprogmodeller kan generere tekst, der er kønsstereotypisk eller diskriminerende. For eksempel kan en sprogmodel være mere tilbøjelig til at generere tekst, der beskriver en kvinde som hjemmegående eller en mand som leder.
- Race: Sprogmodeller kan generere tekst, der er racestereotypisk eller diskriminerende. For eksempel kan en sprogmodel være mere tilbøjelig til at generere tekst, der beskriver en sort person som kriminel eller en hvid person som intelligent.
- Religion: Sprogmodeller kan generere tekst, der er religionsstereotypisk eller diskriminerende. For eksempel kan en sprogmodel være mere tilbøjelig til at generere tekst, der beskriver en muslim som terrorist eller en kristen som hellig.
- Politiske holdninger: Sprogmodeller kan generere tekst, der er politisk stereotypisk eller diskriminerende. For eksempel kan en sprogmodel være mere tilbøjelig til at generere tekst, der støtter et bestemt politisk parti eller ideologi.
- Alder: Sprogmodeller kan generere tekst, der er aldersstereotypisk eller diskriminerende. For eksempel kan en sprogmodel være mere tilbøjelig til at generere tekst, der beskriver en ældre person som svag eller en ung person som uansvarlig.





cfu
københavns professionshøjskole

Her er nogle gode forslag til at arbejde med bias og sprogmodeller i skolen

Start med at introducere eleverne til begrebet bias. Hvad er bias? Hvordan kan bias påvirke vores tanker og handlinger? Hvordan kan bias påvirke sprogmodeller?

- Udforsk eksempler på bias i sprogmodeller. Der findes mange eksempler på bias i sprogmodeller. Eleverne kan arbejde med at identificere og forstå disse eksempler.
- Tal med eleverne om, hvordan bias kan reduceres i sprogmodeller. Der findes en række metoder til at reducere bias i sprogmodeller. Eleverne kan diskutere disse metoder og komme med deres egne ideer.

Et par tips

- **Brug et alderssvarende sprog og eksempler.**
- **Vær åben og ærlig om de udfordringer, der er forbundet med bias i sprogmodeller.**
- **Fremhæv elevernes egen rolle i at reducere bias.**

Ved at arbejde med bias og sprogmodeller i folkeskolen kan vi hjælpe eleverne med at udvikle en kritisk sans og forstå, hvordan bias kan påvirke vores samfund. Arbejde med bias og sprogmodeller kan være en værdifuld læringsoplevelse, da det hjælper elever og studerende med at forstå, identificere og tackle de udfordringer, der er forbundet med AI og automatiserede systemer.
Det kan være i scenarier som:

Etik og ansvar i AI-udvikling
Kildekritik og identifikation af bias i tekster
Bias og samfundsproblemer
Bias i nyheder og information

Det er også vigtigt at understrege, at bias er et komplekst problem, der ikke har en enkel løsning.

cfu
københavns professionshøjskole