

Hvem ville ChatGPT stemme på, og hvorfor er det vigtigt?

CAISA Forskningsbrief

Forfatter

Brandl, Stephanie

Redaktører

Søgaard, Anders; Feldt, Johannes N.

Oversat af

Poulsen, Josefine L.; Feldt, Johannes N.

- Publikationsdato 16. april 2026
- Publiceret af CAISA, Det Nationale Center for AI i Samfundet, København og Aalborg, Danmark
- Ophavsret © Forfatteren 2026
- ISSN 2795-0646
- Dokumentversion Udgivers PDF, registreret version
- Citation for publiceret version (APA) Brandl, S. (2026). Who Would ChatGPT Vote for, and Why Should We Care? *CAISA - Brief*. <https://caisa.dk/en/forskning/hvem-ville-chatgpt-stemme-paa-og-hvorfor-er-det-vigtigt>

Hvem ville ChatGPT stemme på, og hvorfor er det vigtigt?

Af Stephanie Brandl

Resumé

Dette forskningsbrief præsenterer et overblik over vores nuværende viden om politisk bias i store sprogmodeller (LLM'er), og om hvordan de kan påvirke og influere borgere, når de tyer til chatbots for vælgerrådgivning med fokus på danske folketingsvalg. Vi tester forskellige LLM'ers politiske ideologi, deres viden om det danske partisystem, samt hvorvidt de favoriserer bestemte partier over andre, når de anbefaler, hvem man skal stemme på baseret på forskellige vælgerprofiler. Vi bekræfter tidligere studiers resultater og viser, at LLM'er politisk flugter med centristiske partier (Moderaterne, Radikale Venstre, Alternativet, Socialdemokratiet), og at LLM'er, baseret på kandidaters besvarelser, disproportionalt anbefaler bestemte partier (Moderaterne, Liberal Alliance, Dansk Folkeparti og Enhedslisten). Briefets formål er at sætte fokus på, at chatbots ikke bør anses som en pålidelig kilde til vælgerrådgivning i lyset af det danske folketingsvalg i marts 2026 – det første folketingsvalg i Danmark efter lanceringen af ChatGPT i November 2022. Et survey udarbejdet af Digital Democracy Centre viser, at især unge vælgere kan have benyttet chatbots til at træffe deres beslutning om, hvem de skulle stemme på. Vi argumenterer for, at dette kan være problematisk, når det kommer til informationskvalitet, demokratisk deltagelse og digital kritisk tænkning.

Baggrund

Store sprogmodeller, teknologien bag ChatGPT og andre chatbots, nedarver beviseligt sociale bias fra deres

træningsdata, et problem der grundigt studeres, men som endnu ikke er løst. Selvom moderne chatbots er trænet til at følge specifikke retningslinjer, der skal gøre dem hjælpsomme og ikke skadelige, viser tidligere studier på tværs af en række lande og lovgivningssammenhænge, at chatbots indeholder *politiske bias*, der gør det mere sandsynligt, at de anbefaler eller retter sig efter ét parti fremfor et andet.

Disse bias er ikke nødvendigvis tydelige ved første øjekast. Hvis du spørger en chatbot, vil den højst sandsynligt ikke 'fortælle' dig hvem du skal stemme på, eller hvad den 'mener' om en bestemt politik. Dette er et resultat af retningslinjer, der er implementeret som en del af træningsprocessen. Men de mere subtile bias kan være mere misvisende, idet brugere måske antager, at de modtager objektiv information, når dette ikke er tilfældet.

Endnu en årsag til at være forsigtig er politisk *misinformation*.¹ Eftersom de fleste chatbots er trænet af ikke-danske virksomheder er det uklart hvilken – hvis overhovedet nogen – type viden de besidder om det danske politiske partisystem.

I det følgende giver vi baggrundsinformation angående disse anliggender og viser resultaterne fra vores egen biasanalyse. Til slut giver vi en række anbefalinger til videre handling.

Hvorfor har chatbots socialt bias?

LLM'er er trænet på gigantiske tekstdatasæt, der

¹ Misinformation refererer til falsk, ukorrekt eller misledende information der deles uden overlæg, hvilket er forskelligt fra disinformation, som kræver intention, og som derfor er overlagt deling af falsk information.

indeholder data fra hele internettet inklusiv onlinefora, encyklopædier, sociale medier, nyhedsartikler, men også data fra bøger og en række andre kilder. Udover små, samt mere akademiske, initiativer frigiver udviklere sjældent udførlige detaljer angående træningskorporaenes komposition. Fordi LLM'er ikke besidder nogen form for forudgående viden om verden, bygger de til fulde på disse datasæt, hvilket betyder, at det verdensbillede dataen repræsenterer vil blive indbygget i modellen.

I et eftertræningsstadium er samtalebaserede LLM'er, såsom ChatGPT og Claude, yderligere trænet, for at de kan blive hjælpsomme fremfor skadelige. Denne *alignment*-procedure har været genstand for bredere diskussioner, om hvem der definerer, hvad der betragtes som hjælpsomt eller skadeligt, og hvordan retningslinjer virker i praksis. Også her er det meget sjældent, at træningsprotokoller offentliggøres (Chalkidis, 2025).

Både træningskorpuset og træningsprocedurerne kan føre til (og gør det ofte) at LLM'er nedarver bestemte verdenssyn. Sociale bias såsom kønsbias eller racielle bias har været studeret grundigt, men man har ikke fundet en løsning, der kan fjerne dem fuldstændigt. Foregående forskning har også vist, at ChatGPT kulturelt stemmer mere overens med USA end med Kina eller Spanien, især når den promptes på engelsk (Cao et al., 2023).

Ud over de ovenfor beskrevne bias er det ligeledes dokumenteret, at sprogmodeller arver politiske bias. Ét studie evaluerer de politiske holdninger hos tidligere sprogmodeller såsom BERT og RoBERTa, der overvejende er trænet på bogkorpora. Ved anvendelse af det politiske kompas placeres disse modeller økonomisk omkring centrum og svagt traditionalistisk på den sociale akse. Til sammenligning viser studiet, at (dengang) nyere LLM'er, såsom GPT3 og GPT4, som i højere grad er trænet på webbaserede tekster, placerer sig økonomisk på centrum-venstre og let til venstre på den sociale akse. Forfatterne fremsætter den hypotese, at karakteren af træningskorpora spiller en væsentlig rolle i denne forskel: Webtekster har tendens til at være mere liberale/libertarianske end (ældre) bogtekster, og den efterfølgende *alignment*-proces blev anvendt for GPT3/4, men ikke for BERT-baserede modeller (Feng et al., 2023).

Hvordan påvirker disse bias brugere?

Når først modellerne har nedarvet sociale bias kan brugerne påvirkes i anvendelsens forskellige stadier. Det

kan ske, uden de ved det, når en LLM bruges 'i baggrunden' til at rådgive om sundhedsbehandling, juridiske afgørelser eller ansættelsesbeslutninger, hvor forskellige bagvedliggende karakteristika kan medføre ufordelagtige situationer hvor kvinder fremstår mindre kvalificerede til stillinger i mandsdominerede erhverv, eller hvor racialiserede borgere får ringere muligheder for at blive frisat frem mod deres retssag. Samfundsmæssige bias kan også påvirke en LLM's præstation i direkte brug, når en bruger spørger efter råd eller anbefalinger til, f.eks., hvem de skal stemme på. Modellen kan komme til at misrepræsentere et parti med hensyn til dets politiske agenda eller anbefale et parti der ikke passer til brugerens profil.

Hvad har dette at gøre med danske valg?

Selvom det er muligt at kommunikere flydende på dansk med de førende chatbots såsom Google's Gemini eller ChatGPT, kan vi antage, at størstedelen af den træningsdata modellerne har set, kommer fra engelsksprogede kilder i en nordamerikansk kontekst. Det kan betyde, at modellerne har begrænset viden om dansk kultur og det danske politiske landskab. Samtidig har en spørgeskemaundersøgelse vist, at mange borgere tyer til chatbots for råd til, hvad de skal stemme. Hvor 20% af danske vælgere i 2026 i alderen 18 til 28 år forventes at anvende AI, i kontrast til 5% i aldersgruppen 61-79 år (DDC, 2026). Ufuldstændig 'viden' plus politisk bias fører potentielt til misinformation gennem LLM'er og skævhed i stemmeanbefalingerne, hvilket gør dem uegnede til at være pålidelige stemmerådgivere på nuværende tidspunkt.

Analyse

Hvordan positionerer LLM'er sig i Kandidattesten?

For at identificere politiske bias i de førende chatbots på tværs af det danske politiske landskab anvender vi Altingets Kandidattest fra folketingsvalget i 2022 og sammenligner modellernes svar med de daværende kandidaters svar. Til dette formål klassificerer vi udsagnene i henhold til det politiske kompas langs henholdsvis den økonomiske akse (venstre-højre) og den sociale akse (libertariansk-traditionalistisk) med henblik på at placere både de politiske partier og modellerne inden for disse dimensioner. Vi tester både open-weight-modeller (Llama, Qwen, Gemma og Mistral) og proprietære modeller (Gemini samt GPT-3/4/5).²

Samlet set finder vi, at LLM'er er positioneret inden for et centrum-venstre/libertariansk spektrum (se Appendix). I

² For både modellernes og kandidaternes svar multiplicerer vi graden af enighed (fuld uenighed/enighed: -1/1; delvis uenig/enig: -0,5/0,5) med udsagnenes koordinater på det politiske kompas. Herefter beregnes gennemsnittet på tværs af samtlige svar for at opnå de endelige koordinater i det todimensionelle økonomisk/sociale koordinatsystem.

relation til de danske partier svarer denne placering omtrent til et kontinuum mellem Socialdemokratiet og Alternativet. Disse resultater er i overensstemmelse med andre studier – ud over det beskrevet ovenfor – der analyserer dialogbaserede LLM'er på tværs af flere lande, herunder Tyskland, Norge, Nederlandene, Spanien samt EU (Chalkidis & Brandl, 2024, Rettenberger et al., 2025, Chen et al., 2026).

Hvordan placerer LLM'er danske partier i Kandidattesten?

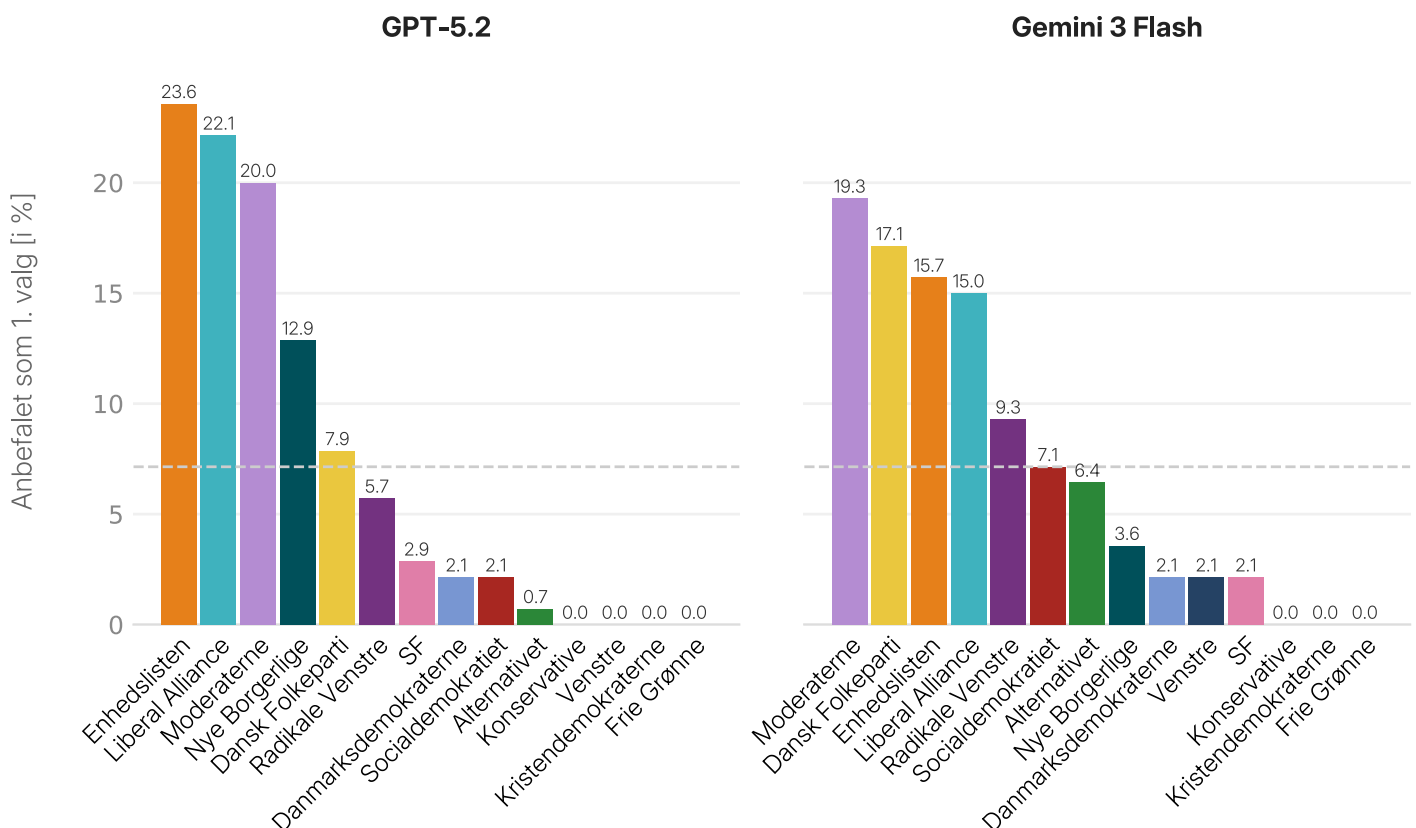
Endvidere promptede vi samtlige modeller til at besvare Kandidattesten hvor vi påtog os rollen som medlem af et specifikt politisk parti med henblik på at undersøge, hvor præcist LLM'erne er i stand til at placere danske partier. Denne test blev gennemført for Enhedslisten, Socialdemokratiet og Nye Borgerlige for at dække hele det politiske spektrum samt det primære regeringsbærende parti. Nyere versioner af proprietære modeller såsom GPT-5.2 og Gemini-3-Flash placerer alle tre partier meget tæt på deres faktiske positioner på det todimensionelle politiske kompas. Ydermere observerer vi, at tidligere versioner – herunder GPT-3.5 samt Qwen og Mistral – tenderer til at placere de tre partier meget tæt på hinanden i en centrum-venstre/libertariansk position. Da denne test fungerer som en indikator for modellernes viden om

danske politiske partier, er det bemærkelsesværdigt, at Qwen og Mistral er de eneste modeller, der er trænet af virksomheder med hjemsted uden for USA (henholdsvis Kina og Frankrig).

Hvilke partier anbefaler LLM'er at stemme på?

Et studie i forbindelse med det hollandske valg i 2025 viste, at LLM'er anbefaler visse politiske partier hyppigere end andre, selv når de præsenteres for et lige stort antal vælgerprofiler for hvert parti (RAN, 2025). Vi replikerede dette studie ved at instruere GPT-5.2 og Gemini-3-Flash med vælgerprofiler baseret på den danske Kandidattest fra 2022. Ligesom i det hollandske studie anvendte vi de svar, som partiernes kandidater havde afgivet, til at sammensætte et lige stort antal syntetiske vælgerprofiler pr. parti. For at gøre profilerne mere realistiske ændrede vi tilfældigt omkring 20% af svarene pr. profil. Herefter gennemførte vi 10 iterationer for hver model og hvert parti og bad modellerne om at returnere en liste med tre partier rangeret efter relevans. Vi finder, at begge modeller i uforholdsmæssig grad anbefaler visse partier oftere end andre. I figur 1 præsenterer vi rangordningen baseret på partiernes placering som førstevalg samt den tilhørende baseline-sandsynlighed, dvs. den andel, hvormed hvert parti var repræsenteret i datasættet.³ Resultaterne viser, at

Figur 1: Stemmeanbefalinger fra to LLM'er.



LLM'er blev instrueret med et lige stort antal vælgerprofiler for hvert parti, baseret på Kandidattesten 2022. Modellerne blev bedt om at anbefale et passende parti. Søjlerne viser fordelingen af partianbefalinger.

³ Af hensyn til læsbarheden har vi i figuren og den tilhørende tekst anvendt forkortede betegnelser for partierne, mens vi i de oprindelige prompts benyttede partiernes fulde navne. Analysen omfatter udelukkende de partier, for hvilke der forelå svar til Altingets Kandidattest i 2022.

GPT tydeligt favoriserer bestemte partier frem for andre. Enhedslisten, Liberal Alliance og Moderaterne udgør tilsammen 65% af samtlige førsteprioritetsanbefalinger, sammenlignet med en baseline-sandsynlighed på omkring 21%, hvis alle partier blev anbefalet ligeligt, dvs. proportionalt med deres forekomst. Omvendt blev Konservative, Venstre, Danmarksdemokraterne og Frie Grønne slet ikke anbefalet som førstevalg. Geminis anbefalinger er en smule mere ligeligt fordelt, men modellen anbefaler ligeledes partier såsom Moderaterne, Dansk Folkeparti og Enhedslisten væsentligt oftere, end de tilsvarende vælgerprofiler blev præsenteret, samtidigt med at Konservative, Kristendemokraterne og Frie Grønne helt udelukkes. Denne skævhed kan ikke forklares ved partirelaterede faktorer såsom ideologi, medlemstal, parlamentarisk repræsentation eller partiernes alder. Blandt de både over- og underrepræsenterede partier findes således både partier fra venstre- og højresiden af det politiske spektrum samt både nyere og veletablerede partier.

Det vi ser er, at modeller ikke bevæger sig langt væk fra det politiske spektrum, men snarere at de samler sig om specifikke partier til venstre (Enhedslisten), i centrum (Moderaterne) og til højre (Liberal Alliance/Dansk Folkeparti/Nye Borgerlige). Det er et fund, der til dels flugter med det studie, der blev udført i Holland, hvor partianbefalinger for det meste tilfaldt ét venstrefløjsparti og ét højrefløjsparti, men intet midterparti.

Hovedpointer og implikationer

Misinformation og skævvredne stemmeanbefalinger udgør en risiko ved anvendelse af chatbots til politisk stemmerådgivning. Samtidigt er det alment kendt, at særligt unge mennesker i stigende grad benytter chatbots, når de søger råd i deres dagligdag, hvilket med stor sandsynlighed også omfatter politisk rådgivning. Nedenfor sammenfatter vi vores væsentligste resultater og anbefalinger:

- Vores resultater, sammenholdt med tidligere forskning, indikerer, at nuværende LLM'er ikke er egnede som stemmerådgivere. Vi viser, at kun et begrænset antal LLM'er har en nogenlunde forståelse af danske politiske partier. Endvidere finder vi, at samtlige LLM'er befinder sig centralt på det politiske spektrum, hvilket bekræfter tidligere studier fra andre europæiske lande samt USA.
- I betragtning af fraværet af en entydig definition af politisk neutralitet er det vanskeligt fuldstændigt at fjerne politisk bias eller blot at definere, hvordan en politisk neutral model bør udformes. I modsætning til bias relateret til eksempelvis køn eller race, hvor klare

lighedsmaal kan formuleres, gør dette sig ikke gældende for politisk bias.

- Yderligere evaluering viser, at to af de mest anvendte LLM'er – modellerne bag Google Gemini og ChatGPT – favoriserer visse politiske partier frem for andre, selv når de instrueres med et lige stort antal vælgerprofiler på tværs af alle partier. Indtil videre kan der ikke identificeres et entydigt mønster i forhold til partiernes ideologi, parlamentariske repræsentation eller graden af institutionel etablering.
- De præsenterede resultater er baseret på kontrollerede eksperimenter. Der er derfor behov for yderligere undersøgelser i form af brugerundersøgelser, hvor deltagere indgår i realistiske interaktioner med henblik på at modtage politisk rådgivning, der tager udgangspunkt i deres interesser og individuelle profiler.
- På længere sigt er der behov for mere intuitive, fleksible og brugervenlige alternativer til de nuværende kandidattests, hvor udsagn er foruddefinerede, ofte svære at forstå og ikke nødvendigvis afspejler vælgernes faktiske interesser. Udviklingen og træningen af en sikker og pålidelig chatbot kan i denne sammenhæng udgøre en relevant og perspektivrig mulighed.

Om forfatteren

Stephanie Brandl er adjunkt ved Center for Social Data Science på Københavns Universitet. Hun gennemførte dette projekt under sit to måneder lange fellowship ved CAISA, Det Nationale Center for AI i Samfundet i februar-marts 2026.

Om CAISA

Det Nationale Center for Kunstig Intelligens i Samfundet (CAISA) er et nationalt konsortium, der samler forskere fra Københavns Universitet, Aalborg Universitet, Aarhus Universitet, IT-Universitetet og DTU i tæt samarbejde med Pioneer Centre for AI (P1).

Som Danmarks uafhængige forskningscenter for kunstig intelligens i samfundet sætter CAISA borgerne i centrum. Vi udfører banebrydende tværfaglig forskning og skaber overblik over nye videnskabelige gennembrud. Funderet i ny og tværfaglig forskning rådgiver vi beslutningstagere i den offentlige og private sektor i, hvordan de bedst udvikler og anvender kunstig intelligens i praksis, så den bidrager til vækst, understøtter demokratiet og styrker digital selvbestemmelse.

Om CAISAs briefs

CAISAs briefs er en del af CAISAs indsats for at sikre, at viden og nye indsigter fra forskningsverdenen styrker beslutningstagere i offentlige myndigheder og private virksomheder – og dermed samfundet som helhed – når det står overfor de muligheder og risici, som hastig teknologisk forandring medfører. CAISA udgiver to slags briefs:

Forskningsbriefs præsenterer forsknings- og evidensbaseret viden inden for AI og samfund i en tilgængelig form.

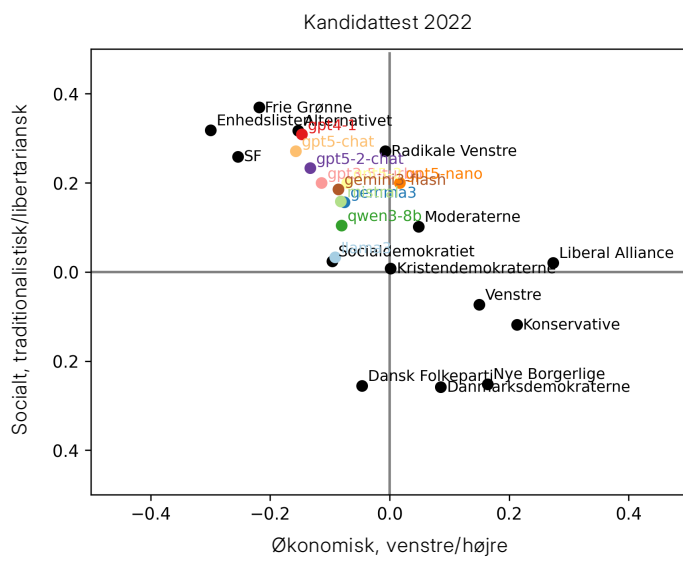
Positionsbriefs udtrykker forfatterens forskningsbaserede og informerede vurdering af vigtige problemstillinger relateret til AI og samfund.

CAISAs briefs udgives under redaktion af Anders Søgaard, der er professor ved Datalogisk Institut, Københavns Universitet og chefforsker i CAISA, samt Johannes N. Feldt, der er videnskabelig assistent i CAISA. Alle briefs læses af og modtager kommentarer fra mindst én ekstern uafhængig forsker inden udgivelse.

Forfatterne er ansvarlige for indholdet i et CAISA-brief.

Appendix

Figur 2: Politisk kompas der viser, hvordan modellerne placerer sig selv i Kandidattesten fra 2022.



Referencer

- Cao, Y., Zhou, L., Lee, S., Piqueras, L. C., Chen, M., & Hershcovich, D. (2023, May). Assessing cross-cultural alignment between ChatGPT and human societies: An empirical study. In Proceedings of the first workshop on cross-cultural considerations in NLP (C3NLP) (pp. 53-67).
- Chalkidis, I., & Brandl, S. (2024). Llama meets EU: Investigating the European political spectrum through the lens of LLMs. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers) (pp. 481-498).
- Chalkidis, I. (2025). Decoding alignment: A critical survey of llm development initiatives through value-setting and data-centric lens. arXiv preprint arXiv:2508.16982.
- Chen, J., de Jong, K., Poole, A., Burakowski, J., Nosti, E. E., Windt, J., & Wang, C. (2026). Uncovering Political Bias in Large Language Models using Parliamentary Voting Records. arXiv preprint arXiv:2601.08785.
- Feng, S., Park, C. Y., Liu, Y., & Tsvetkov, Y. (2023, July). From pretraining data to language models to downstream tasks: Tracking the trails of political biases leading to unfair NLP models. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 11737-11762).
- Digital Democracy Centre [@ddc-sdu.bsky.social] (2026, March 20). DDC-SURVEY: DANSKERNE VIL BRUGE CHATBOTS TIL STEMMERÅD [...] [Post]. Bluesky. <https://bsky.app/profile/ddc-sdu.bsky.social/post/3mhih7sildk2z>
- Report AI & Algorithms Netherlands, Department for the Coordination of Algorithmic oversight. (2025). AI-chatbots as voting aid. <https://www.autoriteitpersoonsgegevens.nl/en/system/files?file=2025-10/AI%20chatbots%20as%20voting%20aid.pdf>
- Rettenberger, L., Reischl, M., & Schutera, M. (2025). Assessing political bias in large language models. Journal of Computational Social Science, 8(2), 42.



CAISAs hjemmeside



CAISAs LinkedIn