caisa Det Nationale Center for AI i Samfundet
The National Center for AI in Society

Who would ChatGPT vote for, and why should we care?

CAISA Research Brief

Author
Brandi, Stephanie

Editors
Søgaard, Anders; Feldt, Johannes N.

- Publication date: 16 April 2026
- Published by: CAISA, The National Center for AI in Society, Copenhagen and Aalborg, Denmark

- ISSN: 2795-0646
- Document version: Publisher's PDF, registered version
- Quotation for published 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

Who would ChatGPT vote for, and why should we care?

By Stephanie Brandl

Summary

This research brief presents an overview of our current knowledge about political bias in large language models (LLMs) and how this can affect and influence citizens when they turn to chatbots for voting advice, focusing hereby on Danish elections. We test various LLMs on their political ideology, their knowledge about the Danish party systems and whether they favour certain parties over others when recommending who to vote for based on provided voter profiles. We show that LLMs on policy issues align with centrist parties (Moderaterne, Radikale Venstre, Alternativet, Socialdemokratiet), and that LLMs based on candidate responses disproportionally recommend specific parties (Moderaterne, Liberal Alliance, Dansk Folkeparti, and Enhedslisten). The purpose of this brief is to raise awareness that chatbots should not be considered a reliable source for voting advice in light of the Danish parliamentary election in March 2026, the first national election in Denmark after the release of ChatGPT in November 2022. Based on a survey by the Digital Democracy Centre, in particular young voters might have turned to chatbots when making their decision on who to vote for. We argue that this might be problematic with respect to information quality, democratic participation, and digital critical thinking.

Background

Large language models (LLMs), the technology behind ChatGPT and other chatbots, are known to inherit social biases from their training data, a topic widely studied yet not solved. Although modern chatbots were usually trained to follow specific guardrails to be helpful and not harmful, prior work across multiple countries and legislations shows that chatbots contain *political biases* that make them more likely to recommend or align with one party over another.

These biases might not be obvious at first sight. If you ask a chatbot, it will most likely not "tell" you who to vote for or what it "thinks" about a certain policy. This is a result of guardrails implemented as a part of the training process. However, the more subtle biases could be even more misleading as users might assume to receive objective information when in fact, they are not.

Another reason to be cautious is potential *misinformation*.[1] Since most chatbots were trained by non-Danish companies, it is unclear what, if any, type of knowledge they have about the Danish political party system.

In the following, we will give some background information on those concerns and show results on our own bias analysis. In the end, we will give some recommendations on how to proceed from here.

Why are chatbots socially biased?

LLMs are trained on massive text datasets that contain data from the entire web, including online forums, encyclopaedias, social media, newspaper articles but also data from books and many other sources. Apart from often smaller and more academic initiatives, developers do not

[1] Misinformation refers to false, inaccurate or misleading information that is shared unintentionally which is different from disinformation which requires intent and thus a deliberate sharing of false information.

disclose full details on the composition of these training corpora. As LLMs do not have any prior knowledge of the world, they fully rely on these datasets which means that whatever world view they represent will be imposed onto the model.

In a post-training stage, conversational LLMs, such as ChatGPT or Claude, are then further trained in order to be helpful and not harmful. This alignment procedure has been subject to a broader discussion as to who defines what is considered helpful or harmful and how well those guardrails work in practice. Also here, details on training protocols are very rarely published (Chalkidis, 2025).

Both the training corpus as well as training procedures can, and often do, lead to LLMs inheriting certain world views. Social biases such as gender bias and racial biases are widely studied, yet no solution has been found on how to remove them completely. Previous work has also found that ChatGPT aligns culturally more with the US than with China or Spain, especially when prompted in English (Cao et al., 2023).

Besides the aforementioned biases, it has also been shown that models inherit political biases. One study evaluates political leanings of earlier language models such as BERT and RoBERTa which were predominantly trained on book corpora. The authors use the political compass and find them to be positioned economically in the centre and slightly traditionalist on the social axis. Whereas (at the time) newer, LLMs such as GPT3 and GPT4 whose training data included more web data are found economically on the centre-left and slightly libertarian on the social axis. The authors hypothesize that the nature of the training corpora played an important role here: web texts tend to be more liberal/libertarian than (older) book texts, as well as the additional post-training alignment procedure that was carried out for GPT3/4 but not for BERT-based models (Feng et al., 2023).

How do these biases affect users?
Once the model inherits social biases, this can affect users at different stages of deployment. It might happen without their knowledge when an LLM is used "in the background" to provide advice on health treatments, legal rulings or hiring decisions where different protected attributes might lead to disadvantages such as women being considered less qualified for male-dominated jobs or racialized citizens facing a lower chance for getting out on bail until their trial. Societal biases can also influence an LLM's performance in direct usage when a user is asking for advice or recommendations such as who to vote for. The model might misrepresent a party with respect to their political agenda or recommend a party that does not fit the user's profile.

What does this have to do with Danish elections?
Although it is possible to communicate fluently with state-of-the-art chatbots such as Google's Gemini or ChatGPT in Danish, we can assume that most training data the models have seen came from English sources in a North American context. This might lead to incomplete knowledge about Danish culture and the Danish political landscape. At the same time, a survey finds that many citizens turn to chatbots for voting advice, where 20% of Danish voters in 2026 between 18 and 28 are expected to use AI in contrast to 5% of the age group 61-79 (DDC, 2026). Incomplete "knowledge" plus political bias potentially leads to misinformation through LLMs and skewed voting recommendations which make them unsuitable to be reliable voting advisors at this point. We investigate this in more detail.

Analysis

How do LLMs place themselves in the Kandidattest?
To identify political biases in state-of-the-art chatbots across the Danish political landscape, we use Altinget's Kandidattest from the previous national election in 2022 and compare the models' answers with those of the candidates at the time. For this, we label statements according to the political compass in the economic (left-right) and social (libertarian-traditionalist) axes to place both the parties and the models along these. We test open-weight (Llama, Qwen, Gemma & Mistral) and proprietary (Gemini and GPT 3/4/5) LLMs.[2]

Overall, we find LLMs to be positioned on a centre-left/libertarian spectrum (see Appendix). In relation to Danish parties, this aligns on a spectrum between Socialdemokratiet and Alternativet. These findings are in line with other studies, in addition to the one described above, that test conversational LLMs across multiple countries such as Germany, Norway, the Netherlands, Spain and the EU (Chalkidis & Brandl, 2024, Rettenberger et al., 2025, Chen et al., 2026).

How do LLMs place Danish parties in the Kandidattest?
We further prompted all models to answer the Kandidattest while impersonating a member of a specific party to see

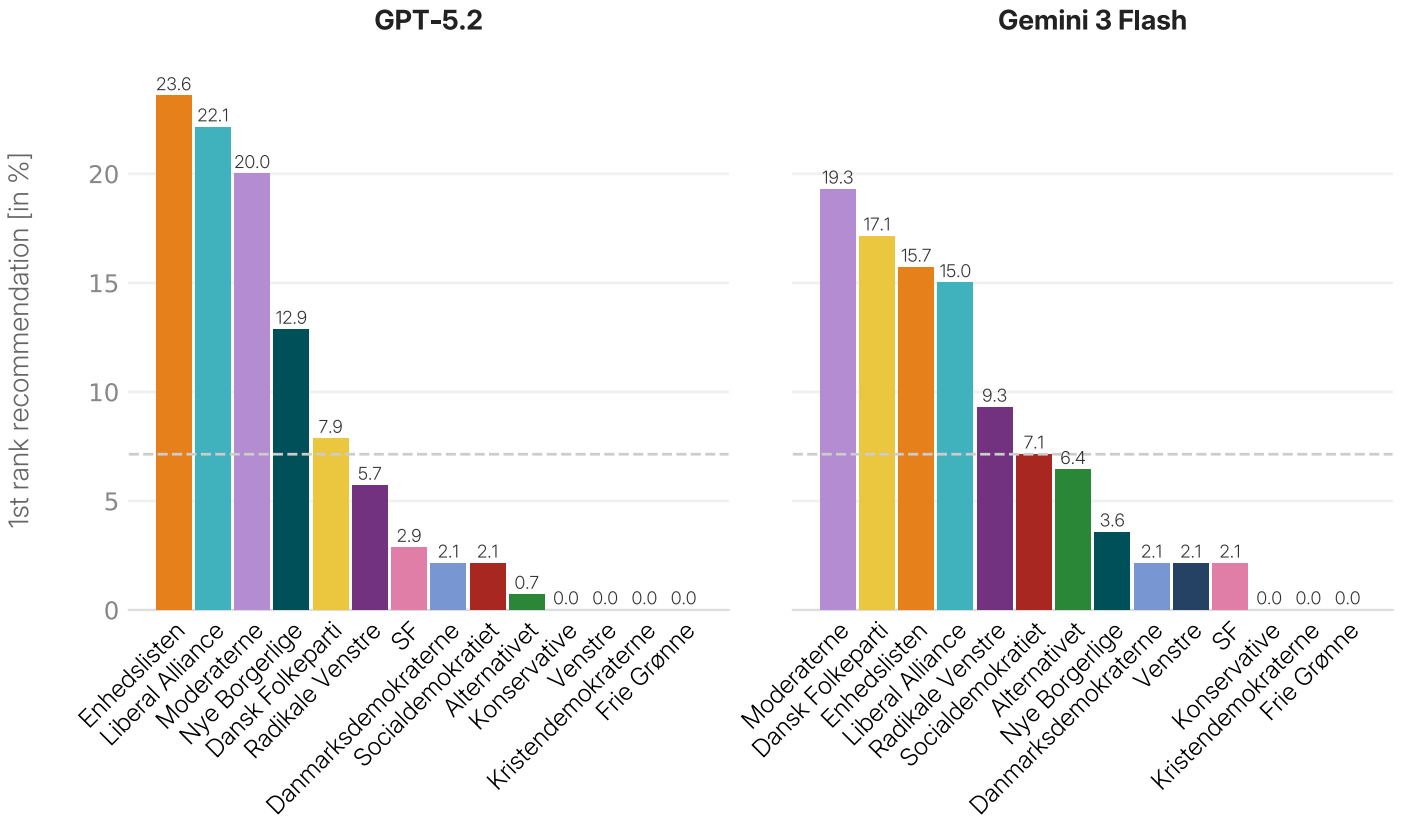
[2] For both, the models' as well as the candidates' answers, we multiply their level of agreement (full dis/agreement: -1/1, slight dis/agreement: -0.5/0.5) with the coordinates of the statements on the political compass and average across all answers to obtain the final coordinates on the two-dimensional economic/social grid.

how accurately the LLMs can place Danish parties. We tested this on Enhedslisten, Socialdemokratiet and Nye Borgerlige to cover the full spectrum and the main governing party. While recent versions of proprietary models such as GPT 5.2 and Gemini 3-flash place all three parties very close to their actual position on the 2-dimensional political compass, we find earlier versions such as GPT 3.5 as well as Qwen and Mistral to place the 3 parties very close together in a centre-left/libertarian position. As this test serves as a proxy for knowledge about Danish parties, it is noteworthy that Qwen and Mistral are the only models that were trained by companies outside the US (China and France, respectively).

Which parties do LLMs recommend voting for?

A study in relation to the Dutch elections in 2025 found that LLMs recommend some parties more often than others when seeing an equal number of voter profiles for each party (RAN, 2025). We replicated their study by prompting GPT-5.2 and Gemini 3 Flash with voter profiles based on the Danish Kandidattest from 2022. As in the Dutch study, we used the answers provided by the party candidates to compose an equal number of synthetic voter profiles per party. We randomly changed around 20% of the answers per profile to make them more realistic. We then ran 10 iterations for each model and party and asked the model to return a list of 3 parties ranked by relevance. We find that both models disproportionally recommend certain parties more often than others. We show the ranking based on the first answer in figure 1 including the baseline probability, i.e., the fraction to which each party was represented.[3] We see that GPT clearly favours some parties over others. Enhedslisten, Liberal Alliance and Moderaterne combined represent 65% of all 1st rank recommendations in comparison to a baseline likelihood of around 21%, if all parties were recommended equally, i.e., according to their appearance. On the other hand, Konservative, Venstre, Danmarksdemokraterne and Frie Grønne were not recommended at all as a first choice. Gemini's recommendations are slightly more equally distributed, but it also recommends parties such as Moderaterne, Dansk Folkeparti and Enhedslisten much more often than respective voter profiles were shown and excludes Konservative, Kristendemokraterne and Frie Grønne. This discrepancy cannot be explained by party ideology, number of members, seats in parliaments or how recently a party was founded. Among the over- and underrepresented parties, there are parties from the left

Figure 1: Voting recommendations from two LLMs.



LLMs were prompted with an equal number of voter profiles for each party based on the Kandidattest 2022. Models were asked to recommend a suitable party. Bars show the distribution over recommendations.

[3] We shortened names for the parties in the figure and the text for readability, in the original prompts we used the full names. We included those parties for which statements to Altinget's Kandidattest in 2022 were provided.

and right political spectrum, as well as newer and established parties. What we see is that recommendations do not move too far away from the political spectrum but rather are conflated towards specific parties on the left (Enhedslisten), centre (Moderaterne) and right (Liberal Alliance/Dansk Folkeparti/Nye Borgerlige). A finding that aligns in parts with the study carried out in the Netherlands where party recommendations mostly fell onto one left-wing and one right-wing party but none on the centre.

Main points and implications

Misinformation and skewed voting recommendations pose a risk when using chatbots for voting advice. At the same time, in particular young people turn to chatbots when seeking advice in their everyday life which most likely also includes political advice. We summarize our main points and recommendations:

- Our results and prior work indicate that current LLMs are not suitable as voting advisors. We show that only a few LLMs have a decent understanding of Danish political parties. We further find that all tested LLMs position themselves on the centre of the political spectrum which confirms previous work in other European countries and the US.
- Considering the lack of a clear definition towards political neutrality, it is challenging to remove political bias completely or even define what a politically unbiased model should look like. As opposed to gender or racial bias where a clear equality objective can be formulated, this is not the case for political bias.
- Additional evaluation shows that two of the most used LLMs, the ones behind Google's Gemini and ChatGPT, favour certain parties over others when being prompted with an equal number of voter profiles across all parties. So far there is no clear pattern visible with respect to party ideology, representation in parliament or how established they are.
- The results we show are based on controlled experiments. This needs to be further investigated in user studies where participants can have realistic interactions in order to get political advice based on their interests and profiles.
- In the long run, we need more intuitive, flexible and user-friendly alternatives to current political questionnaires where statements are pre-selected, often hard to understand and do not necessarily cover voters' interests. Training a safe and reliable chatbot might be a possibility worth pursuing.

About the author

Stephanie Brandl is Assistant Professor at the Center for Social Data Science at the University of Copenhagen She conducted this project during her 2-months fellowship at CAISA in February/March 2026.

About CAISA

The National Center for Artificial Intelligence in Society (CAISA) is a national consortium that gathers researchers from the University of Copenhagen, Aalborg University, Aarhus University, the IT University of Copenhagen and the Technical University of Denmark in close collaboration with the Pioneer Centre for Artificial Intelligence (P1).

As Denmark's independent research center for artificial intelligence in society, CAISA centers the citizen. We carry out groundbreaking interdisciplinary research, and we deliver overview of the most recent scientific breakthroughs. Based on new and interdisciplinary research, we advise decision-makers in the public and private sectors on how they may develop and use artificial intelligence practically, such that it contributes to growth, supports democracy and strengthens digital autonomy.

About CAISA's briefs

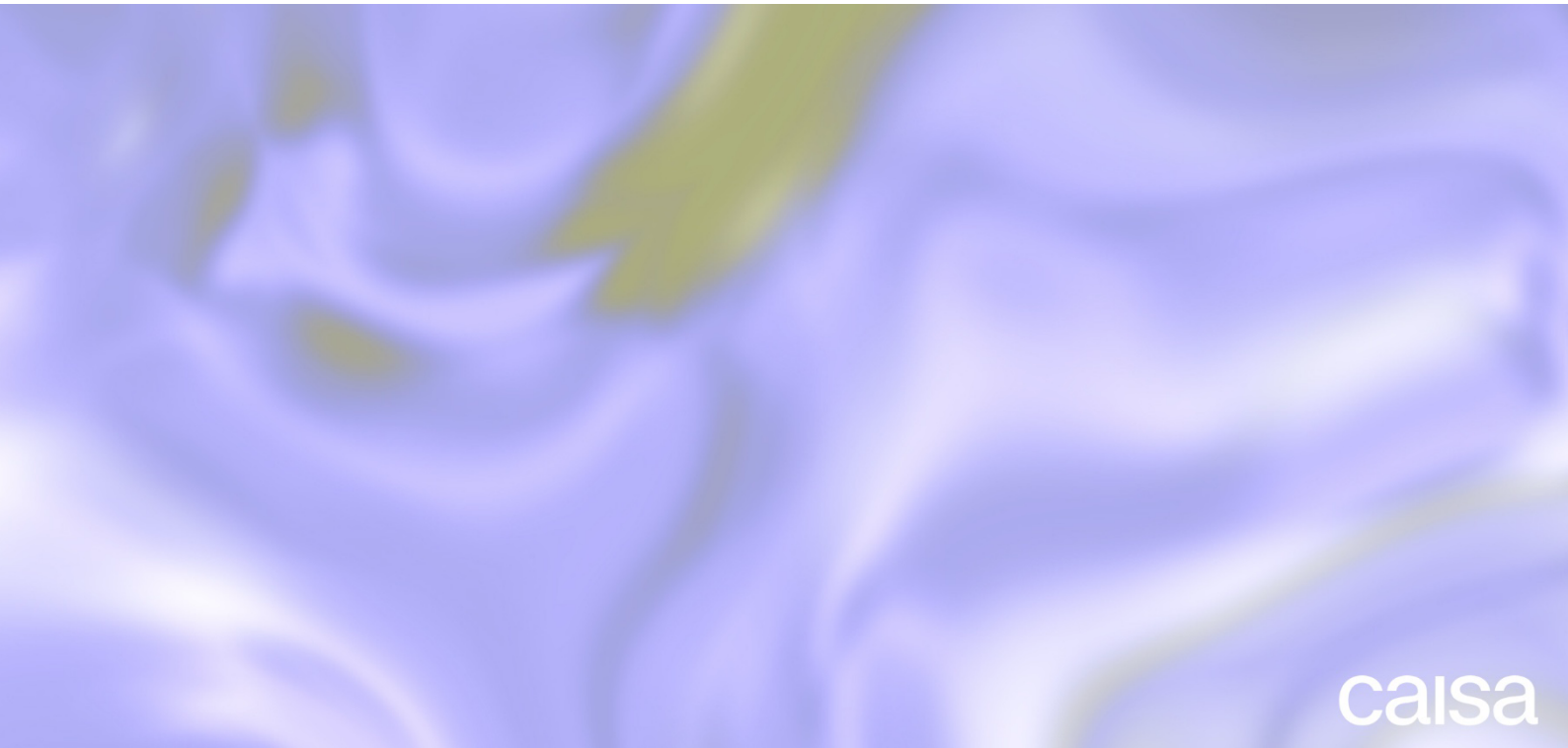
CAISA briefs are part of CAISA's effort to ensure that knowledge and novel insights from within the research community come to empower decision-makers in public and private sectors, and, in turn, society-writ-large when faced with the opportunities and risks posed by rapid technological change. CAISA publishes two types of briefs:

Research briefs present research and evidence-based knowledge about AI and society in an accessible way.

Position briefs express the authors' research based and informed assessment of important challenges related to AI and society.

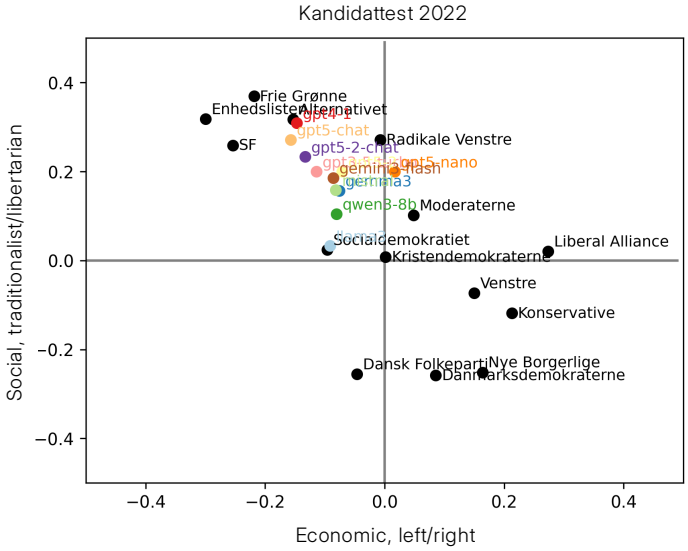
CAISA's briefs are edited by Anders Søgaard, Professor at the Department of Computer Science at the University of Copenhagen and Chief Scientist at CAISA, and Johannes N. Feldt, Research Assistant at CAISA. All briefs are read by – and receive comments from – at least one external independent researcher before publication.

The authors are responsible for the contents of a CAISA brief.



Appendix

Figure 2: Political compass showing how models place themselves in the Kandidattest.



References

- 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.

CAISA's website

CAISA's LinkedIn