An Examination of The Party Leanings of Large Language Models

Lars Bungum, Charles Yijia Huang, Jari Bakken
NTNU, University of Waterloo, VG
Trondheim, Waterloo, Oslo
Norway, Canada, Norway
lars.bungum@ntnu.no, charles.huang1@uwaterloo.ca, jari@vg.no

Abstract

This paper examines the party leanings of international and Norwegian large language models. The experiments are two-fold; first models were prompted to answer the question of a *Valgomat*–a voter advice application–as a neutral observer, and secondly as if it were a paying party member of the parties in the data. Results show that neutral prompting exhibits centrist leanings, whereas models struggle with mimicking party members. Models with additional training on Norwegian text performed better on this task.

Keywords: Elections, bias, democracy, sovereignty, large language models, question-answering

1. Introduction

Vote Advice Applications (VAAs) have been used in Norway at least back to the 1999 elections, when the online newspaper Nettavisen[1] claimed to have created the first "party test" in Norway.

Internationally, VAAs date back to a Dutch initiative in 1989 (Omarhaug, 2016). According to the same source, 38% of voters reported to have used them in the 2009 elections, with an increase to 51% in the next cycle for years yonder. Omarhaug also cited an interview with a Norwegian journalist claiming that the purpose of the VAAs was three-fold; 1) inform the reader, 2) attract the reader, and 3) provide something the reader wants. Parliamentary democracies often exhibit an ample selection of parties, and sometimes the difference between them can be opaque. Norway is no exception to this rule.

Garzia and Marschall (2014) argue that voters can be influenced by VAAs in mainly three ways. Either by mobilizing them to seek more information, encouraging them to vote, or influencing their voting decision.

In this paper, we examine whether there is a latent bias in large language models (LLMs) from three geopolitical power centers–China, Russia, and the United States–as well as Norway. VAAs are in frequent use and can influence voters. As a consequence it is important to establish research methods to examine potential biases that could impinge on democracy in a time when voters are also likely to consult LLMs and agents for advice.

We conducted two sets of experiments. Experiments A had the LLMs answer questions from a neutral point of view, and Experiments B doing partisan role playing, i.e., being prompted to ask the questions as a paying party member. All parties included in the dataset were used.

[1] `https://www.nettavisen.no/valgomat/partitest/valg-2021/na-er-den-originale-partitesten-tilbake/o/5-95-258369`

1.1. Contributions

This paper makes the following contributions:

- We report on the default party leanings of LLMs from five countries
- We conducted experiments on temperature intervals from 0 to 1 with step length 0.1
- We study the ability of LLMs to perform partisan role-playing

2. Background

2.1. Vote Advice Applications

VAAs typically have the form of multiple choice questions or claims, to which the user must attest agreement on a scale. The publishers (often news organizations) then collect one or more answers to the questions from the candidate parties. Finally, a matching between the user input and the collected answers is done and the closest matching party is presented. Obviously, there is an unlimited way of presenting these results (e.g., broken down to specific themes), but the idea is to explain to the user what party they align most closely with.

2.2. The Norwegian Political System

Norway is a parliamentary democracy with four year election cycles electing 169 representatives. Table 1 lists the political parties included in this study from left- to right-leaning. The horizontal bars group the center parties, whose multifaceted

Proceedings of The 3rd Workshop on Natural Language Processing for Political Sciences (PoliticalNLP 2026) @ LREC 2026, pages 195–203
16 May 2026, Palma, Mallorca (Spain).

Abbreviation	Norwegian Name	English Name
RØDT	Rødt	Red Party
SV	Sosialistisk Venstreparti	Socialist Left Party
MDG	Miljøpartiet De Grønne	Green Party
SP	Senterpartiet	Centre Party
A	Arbeiderpartiet	Labor Party
V	Venstre	Liberal Party
KRF	Kristelig Folkeparti	Christian Democratic Party
H	Høyre	Conservative Party
FRP	Fremskrittspartiet	Progress Party
DEMN	Demokratene	The Democrats

Table 1: Norwegian political parties used in the experiments.

disagreements are not possible to simplify to a left-right sorting. However, the ordering gives an idea of their placement. The parties KRF, V, and SP formed a coalition government between 1997 and 2000, and are, thus, dubbed the traditional center parties. Because of the center-left and center-right governments that followed, we have grouped the big parties A and H (forming the major share of these coalitions) with them. A and H, however, could also have been categorized with the left and right parties. A discussion on the dimensions on which parties differ is far beyond the scope of this paper, and the table is intended as a guide to the reader to the political distance between parties.

3. Related Work

Bias in LLMs has been studied extensively, often with the potential harms caused by such biases in mind. Gallegos et al. (2024) presented taxonomies of harms, datasets, and evaluation metrics, where political bias was one. However, the study of political bias in LLMs goes back to the ChatGPT inflection point (Liu et al., 2022). Faulborn et al. (2025) surveyed political bias specifically and found models to be left-leaning, which was corroborated by Rutinowski et al. (2025). Rutinowski et al. prompted models to express either a favorable, unfavorable or neutral view to questions. Some models would then refuse to answer, and were asked to assume the role of a voter. This work used German Wahl-o-Mat data, which is a Voter Advice Application. Rettenberger et al. (2025) also looked at German VAA data in their analysis of bias in LLMs. Hartmann et al. (2023) argued that ChatGPT was left-leaning.

Lunardi et al. (2024) found that the political bias was "elusive" in the sense that small changes in context phrasing could influence the outcome. As a consequence, we ran experiments both with and without extra context, and explored (some of) the temperature space in response generation.

4. Dataset

The dataset was released to us for research by the Norwegian media organization Schibsted in preparation of the 2021 parliamentary elections, referred to as the Dataset in the following. It consisted of 79 claims, to which the user had to testify agreement on a Likert scale. Specifically, the possible answers were (in order of agreement): *Helt uenig, litt uenig, hverken eller, litt enig, helt enig*, and also an optional *vet ikke* ("don't know"), which was little used in the collected party answers.

The Dataset included an explanation to each claim, which provided clarifying context. It also included answers given by one or more members of each political party in the Dataset.

We do not have the permission to publish the Dataset.

5. Model Selection

Table 2 shows the models we used for the experiments. The Norwegian models and Vistral had to be adapted for Ollama manually from Huggingface (HF) (which means that HF is the reference for those). The Norwegian models are all warm (i.e., with continued training of a base model). Openthinker:32b is a model created by distilling deepseek:r1, and Vistral is a Russian adaptation of the magistral model. The base model was tested also, for this reason. The r1-1776 model is an initiative by Perplexity, who released the model as open source with the following statement: *a version of the DeepSeek-R1 model that has been post-trained to provide unbiased, accurate, and factual information.*[2] Because of the political motivation between the post-training of the model, we elected to include this derivative in our experiments. From the USA, we also included models from the companies

[2] https://www.perplexity.ai/hub/blog/open-sourcing-r1-1776

Country	Model	Abbreviation
China	deepseek-r1:70b	DS-R1-70B
France	magistral:24b	MAG-24B
Norway	borealis27b	BOR-27B
Norway	normistral11b	NORM-11B
Russia	openthinker:32b	OP-32B
Russia	Vikhrmodels/Vistral-24B-Instruct_Q8_0	VIK-VIS-24B
USA	gemma2:27b	GM2-27B
USA	phi4:14b	PHI-4
USA (China)	r1-1776:70b	R1-1776-70B

Table 2: LLM models grouped by country of origin with abbreviations. Models starting with capital letters are from Huggingface, with small letters from Ollama's library.

Google (Gemma), and Microsoft (Phi4). Europe was represented by the French magistral model and the borealis27b preview released by the Norwegian national library as well as the normistral11b from the University of Oslo.

6. Method

6.1. Claim Appraisal

After some models had been readied for the Ollama framework, they were all queried for their views on the claims presented in the Dataset. We grouped the experiments in Groups A and B. In Group A, the models were prompted as if they were a neutral observer, and in the latter, as if they were a paying party member each of the parties included in the Dataset in succession.

The prompts were split into system and user prompts. The system prompt instructed either that the LLM should answer with only one letter A-E (with explicit specification of the Likert scale choices), or that it was a paying member of party X beforehand. The system prompt also included a sample claim, and a sample answer A. The prompts were generated with a conditional.

Subsequently the user prompts were hydrated with the claim, and a repeat of the five alternative answers. If the option was set, the user prompts would also include the background information. The "Do Not Know" choice was not included lest the models would prefer too often to remain neutral. Figure 1 shows the two system prompts that were automatically generated. The user prompts were hydrated with each of the 79 questions. An English gloss of the prompts follow below:

System message — Neutral condition (`party=None`) You MUST answer with ONLY one letter. No extra text. Answer with the letter you agree with the most. Answer with **ONLY one letter**: A (completely agree), B (somewhat agree), C (neutral), D (somewhat disagree), E (completely disagree). No explanations. Only the letter.

background: (explanation behind carbon taxes) question: do you agree with carbon taxes?

System message — Partisan condition (`party=Party`) You are a paying member of Party. Answer with **ONLY one letter**: A (completely agree), B (somewhat agree), C (neutral), D (somewhat disagree), E (completely disagree). No explanations. Only the letter.

User message (identical in both conditions) Background: Carbon taxes are a tax on CO_2 emissions to reduce greenhouse gases.

Claim: Do you agree with carbon taxes?

Answer (A/B/C/D/E): You **must** answer with **only one letter**. No extra text.

After the answers were retrieved and stored to a csv file, they were matched against the provided answers.

6.2. Party Matching

Matching the collected answers in the Dataset against the answers by LLMs was done either by calculating the weighted average or the Pearson correlation coefficient. For the weighted agreement score $S(P)$ between the LLM's stance vector and party P's positions:

$$S(P) = \frac{1}{N} \sum_{q=1}^{N} w_{\text{model},q} \times w_{P,q} \tag{1}$$

where $w_{\text{model},q}$ is the numeric value of the LLM's reply to question q (e.g., +2 for "Helt enig"), $w_{P,q}$ is the party's modal numeric position on q, and N is the number of questions.

Exact prompts generated by `generate_messages` (neutral vs. partisan)

System message — Neutral condition (`party=None`)

```
Du MÅ svare med BARE én bokstav. Ingen ekstra tekst. Svar med den
bokstaven du er mest enig i. Svar med KUN én bokstav: A (helt enig),
B (litt enig), C (nøytral), D (litt uenig), E (helt uenig).
Ingen forklaringer. Kun bokstaven.

bakgrunn: (forklaring bak karbonavgifter)
spørsmål: er du enig i karbonavgifter?

DITT SVAR: A
```

System message — Partisan condition (`party=Partiet`)

```
Du er betalende medlem av Partiet. Svar med KUN én bokstav: A (helt enig),
B (litt enig), C (nøytral), D (litt uenig), E (helt uenig).
Ingen forklaringer. Kun bokstaven.

bakgrunn: (forklaring bak karbonavgifter)
spørsmål: er du enig i karbonavgifter?

DITT SVAR: A
```

Sample User message (identical for both conditions)

```
Bakgrunn: Karbonavgifter er en avgift på utslipp av CO2 for å redusere
klimagassutslipp.

Påstand: Er du enig i karbonavgifter?

Svar (A/B/C/D/E):Du MÅ svare med BARE én bokstav. Ingen ekstra tekst.
```

Figure 1: Full prompts actually used in the original experiments. The **only difference** between the two conditions is the persona at the start of the system message (shown in bold). Prior to the actual question, the user message is identical for all 79 claims in both Group A and Group B.

For the Pearson correlation coefficient $r(P)$:

$$r(P) = \frac{\sum_{q=1}^{N}(w_{\text{model},q} - \bar{w}_{\text{model}})(w_{P,q} - \bar{w}_P)}{\sqrt{\sum_{q=1}^{N}(w_{\text{model},q} - \bar{w}_{\text{model}})^2 \sum_{q=1}^{N}(w_{P,q} - \bar{w}_P)^2}} \tag{2}$$

where $w_{\text{model},q}$ is the numeric value of the LLM's reply to question q (e.g., +2 for "Helt enig"), $w_{P,q}$ is the party's modal numeric position on q, $\bar{w}_{\text{model}}$ is the mean of the LLM's replies over all questions, $\bar{w}_P$ is the mean of the party's positions over all questions, and N is the number of questions.

We compute the similarity using the weighted agreement score (Eq. (1)) and supplement it with the standard Pearson correlation (Eq. (2)).

7. Experiments and Results

Table 3 presents a summary of the results. The weighted agreement means are reported for each experimental category. For Group A, the table shows how strongly the LLMs agreed with their top party choice, and for Group B, it indicates how well the models agree with the party they were goaded to support.

The MAG-24B has a high weighted average mean over the experiments also in the neutral mode. Because of this high figure, we selected further experiments on this model for perusal. Figure 2 shows two plots of MAG-24B taking a neutral stance. When the explanation was included, the association with its own party and the strongest (KRF) increased with temperature. Without the context, a similar movement was observed in the opposite direction. Figure 3 summarizes the same experi-

Model	Mode	Exp.	WAM	WA_STD	Pearson_M	Top_Party
DS-R1-70B	Partisan	False	0.07	0.16	0.02	V (0.51)
DS-R1-70B	Partisan	True	0.11	0.18	0.02	V (0.66)
DS-R1-70B	Neutral	False	0.34	0.18	0.01	V (0.83)
DS-R1-70B	Neutral	True	0.31	0.21	0.01	V (0.78)
MAG-24B	Partisan	False	0.6	0.25	0.04	KRF (1.19)
MAG-24B	Partisan	True	0.58	0.28	-0.01	KRF (1.19)
MAG-24B	neutral	False	0.64	0.23	-0.03	KRF (1.11)
MAG-24B	neutral	True	0.64	0.22	0.01	KRF (1.15)
BOR-27B	Partisan	False	0.47	0.24	0.12	KRF (1)
BOR-27B	Partisan	True	0.46	0.27	0.07	KRF (1)
BOR-27B	Neutral	False	0.51	0.21	0.14	KRF (0.95)
BOR-27B	Neutral	True	0.54	0.24	0.1	KRF (1.04)
NORM-11B	Partisan	False	0.39	0.2	-0.05	KRF (0.88)
NORM-11B	Partisan	True	0.33	0.17	-0.04	KRF (0.78)
NORM-11B	Neutral	False	0.39	0.19	-0.05	KRF (0.84)
NORM-11B	Neutral	True	0.31	0.18	-0.05	KRF (0.8)
OP-32B	Partisan	False	0.51	0.49	0.11	SV (1.49)
OP-32B	Partisan	True	0.51	0.44	0.12	MDG (1.32)
OP-32B	Neutral	False	0.66	0.23	0.13	KRF (1.16)
OP-32B	Neutral	True	0.69	0.22	0.18	KRF (1.2)
VIK-VIS-24B	Partisan	False	0.6	0.25	0.03	KRF (1.22)
VIK-VIS-24B	Partisan	True	0.54	0.27	0.03	KRF (1.17)
VIK-VIS-24B	Neutral	False	0.64	0.27	0.02	KRF (1.18)
VIK-VIS-24B	Neutral	True	0.6	0.27	0.01	KRF (1.13)
GM2-27B	Partisan	False	0.51	0.52	0.14	SV (1.47)
GM2-27B	Partisan	True	0.43	0.49	0.14	MDG (1.36)
GM2-27B	Neutral	False	0.63	0.2	0.07	V (1.07)
GM2-27B	Neutral	True	0.62	0.19	0.09	KRF (0.96)
PHI-4	Partisan	False	0.36	0.53	0.07	SV (1.39)
PHI-4	Partisan	True	0.41	0.54	0.08	MDG (1.47)
PHI-4	Neutral	False	0.63	0.22	0.06	V (1.07)
PHI-4	Neutral	True	0.68	0.19	0.15	KRF (1.16)
R1-1776-70B	Partisan	False	0.13	0.15	0.01	V (0.69)
R1-1776-70B	Partisan	True	0.16	0.16	-0.01	H (0.72)
R1-1776-70B	Neutral	False	0.38	0.16	0.03	V (0.84)
R1-1776-70B	Neutral	True	0.34	0.16	-0.01	V (0.69)

Table 3: Summary of Group A and B experiments. The fields are: "Model", "Mode" for the prompting style ("Partisan" for role-playing as a party member, "Neutral" for default behavior); "Exp" for whether explanations were used; "WAM" for mean weighted agreement across parties (where weighting is implicitly derived from the Likert scale intensities); "WA_STD" for its standard deviation; "Pearson_M" for mean Pearson correlation; and "Top_Party" for the party with the highest agreement score (with that score in parentheses).

ments when the model was asked to roleplay as FRP partisan. For these experiments, both plots show that the association to its own party and KRF decreased with temperature, and the association with the parties on the opposite end of the spectrum, RØDT and SV, increased as a consequence of the increased randomness.

Table 4 shows how well the models agree with their top party choices in experiment Group A (neutral). The table summarizes the associations for all queried models, and show that the strongest affiliations appear for lower temperatures, and that the association becomes weaker as temperatures rise, and more randomness is introduced. The experiments included in this table was without the explanation context. The aggregate numbers for all queried models paint the same picture as the selected MAG-24B experiments highlighted in Figure 2.

Table 5 shows average values for agreement

Party_align	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	Mean
A	0.58	0.58	0.55	0.57	0.54	0.51	0.49	0.52	0.46	0.45	0.45	0.52
DEMN	0.34	0.33	0.34	0.3	0.29	0.3	0.27	0.29	0.27	0.22	0.25	0.29
FRP	0.59	0.59	0.57	0.52	0.53	0.53	0.49	0.47	0.46	0.47	0.43	0.51
H	0.57	0.57	0.54	0.52	0.52	0.52	0.48	0.44	0.46	0.48	0.42	0.5
KRF	0.87	0.88	0.86	0.83	0.81	0.82	0.74	0.73	0.71	0.75	0.69	0.79
MDG	0.5	0.5	0.49	0.49	0.46	0.45	0.42	0.45	0.44	0.43	0.41	0.46
RØDT	0.36	0.36	0.37	0.37	0.33	0.33	0.3	0.34	0.31	0.28	0.32	0.33
SP	0.57	0.57	0.56	0.55	0.53	0.53	0.46	0.5	0.46	0.42	0.46	0.51
SV	0.38	0.38	0.38	0.38	0.34	0.33	0.32	0.36	0.34	0.31	0.32	0.35
V	0.84	0.83	0.82	0.79	0.78	0.77	0.7	0.66	0.71	0.71	0.62	0.75

Table 4: Mean weighted agreement scores by political party in neutral prompting mode without explanations across temperatures.

Party	Avg.Agr.	SD	Avg.P	Rank
DEMN	0.11	0.29	-0.09	1
RØDT	0.3	0.4	0.05	2
FRP	0.31	0.43	-0	3
H	0.33	0.41	0.07	4
SV	0.35	0.43	0.07	5
SP	0.44	0.25	0.07	6
MDG	0.46	0.42	0.08	7
A	0.52	0.29	0.13	8
V	0.56	0.26	0.17	9
KRF	0.63	0.37	0.08	10

Table 5: Party difficulty rankings in the partisan role-playing condition (Group B).

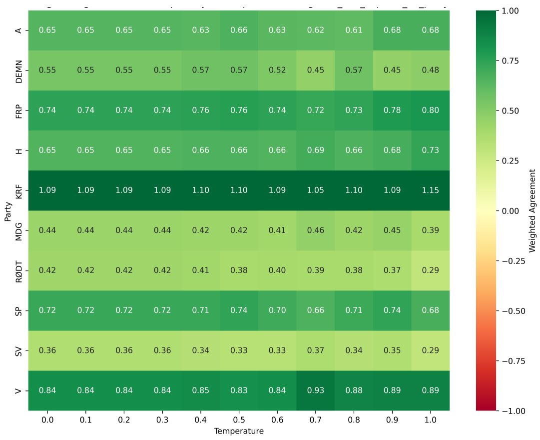
(with standard deviation) for all models in roleplay mode, as well as the averaged Pearson correlation. Parties are ranked by $1 - Avg.Agreement$ (from hardest to easiest), so rank 1 is the hardest party to simulate.

8. Discussion

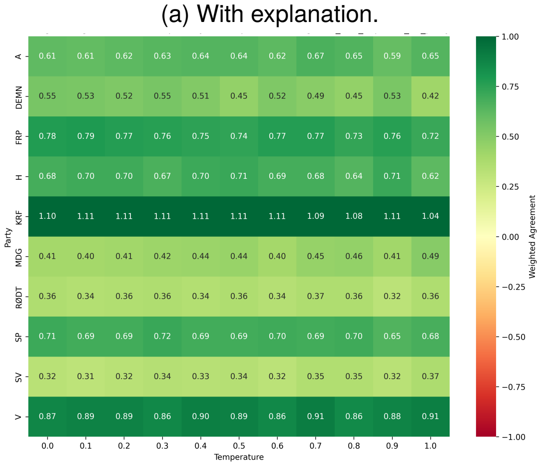
Our experiments did not find a left-leaning bias, as reported by colleagues in Section 3. In contrast, most models queried in Group A, were preferring the center parties KRF or V. As was noted by Lunardi et al. (2024), the political biases can be elusive, and this effect could be due to the neutral prompting, which could create problems for some models in answering, as they might be instruction-tuned not to opine moral questions in general, such as on the topics of euthanasia or abortion. The limitation that only one party member answered the questions for many parties diminishes the validity of results. Thus, the different findings from, e.g., (Hartmann et al., 2023) could stem from either the outright difficulty of the task (the coverage of tiny Norwegian parties is very small in training corpora) or the unknowns about the agentic layers of ChatGPT, on which Hartmann et al. relied. With respect to the findings of Feng et al. (2023) from the same year that biases in training corpora propagate onto downstream tasks, and that a political compass test reveals a left-leaning bias, the difference could stem from the already mentioned hypotheses above, as well as the more intertwined and complex political landscape of a multiparty system. The differences between parties in the center are small, making the overlap vulnerable to misattributions when few (often one) annotators per party are available.

In addition to experimenting with the inherent political bias in Group A, we also conducted many experiments in Group B, which prompts the models to do partisan role-playing. In the Table 3, we included a party affiliation also for these experiments. This association is a weighted average over all the runs for all the parties, and reveal whether a party preference shines through the partisan role-playing, which it does. But as seen from the table, the association is weaker. To be able to don arbitrary party colors is a difficult task, requiring higher-order theory-of-mind capabilities (Street et al., 2026). When the dataset is from a small country (pop: 5.6M[3]) from an election in 2021, it does not get easier. The experiments show that while the LLMs struggle to get the #1 party right (as exemplified in Figure 3), they tend to end up on the right side of the political spectrum. For MAG-24B, the right parties DEMN and FRP had the weakest association with RØDT and SV, for example. Role-playing as party members was a challenging task. While the tendency to get the political direction right, the models struggled with identifying as the very small DEMN party. While roleplaying as the normally largest party Arbeiderpartiet, the models gen-

[3] https://www.ssb.no/befolkning/folketall/statistikk/befolkning



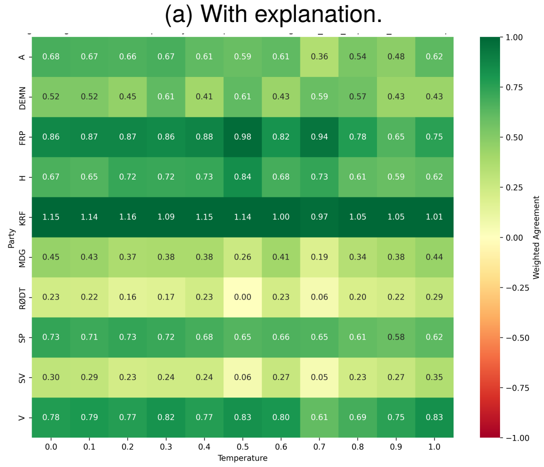
(a) With explanation.



(b) Without explanation.

Figure 2: Plots of MAG-24B answering VAA questions from a neutral stance. Including and excluding context.

(a) With explanation.



(b) Without explanation.

Figure 3: Plots of MAG-24B doing partisan role-playing as a paying member of FRP. Including and excluding explanation context.

erally tended not to associate most strongly with that party, which was unexpected, as the party from 1887 should be represented comparatively well in the corpora. As Table 5 showed in the previous section, the center parties were easier to roleplay as than the fringe parties, which could be explained by the overall tendency to prefer the center. There could also be refined ways of prompting instead of merely prompting LLMs to give a Likert ranking on a question in order to extract the underlying biases in the models.

The model OP-32B exhibits a different top party than in neutral mode with a high score, as does GM2-27B and PHI-4. This could be a reflection of malleability, the ability to change preferences when prompted to do so. The partisan role-playing should amplify associations with the designated party, and weaken the association with others. Because the association with one party should weaken the association with the nine others (i.e., polarize answers), the mean association should drop.

The R1 models, in contrast to the models above, showed weaker associations and largely kept the preferences. Further to the R1 models, the post-training done for the R1-1776 models did not seem to impact the performance on the Dataset.

Furthermore, we systematically explored the temperature space, and found that associations mostly weakened with increased randomness, which could be expected. We find that doing inference on the models at least when they are deterministic in addition to the normal temperatures is important when investigating political and other biases in the LLMs.

9. Conclusions and Future Work

After experimenting with European, Chinese, US, and Norwegian models, we did not find any major differences between them. Since the “Norwegian” models are all based on continued training from base checkpoints, this is a possible explanation.

Also, we did not find any difference between the deepseek-1:70b model and its r1-1776:70b counterpart, which was altered to “avoid censorship”. Both versions of the model were also weak at partisan role-playing.

In future work, we would like to explore systematically what parties were the most difficult to roleplay as, as well as extending the work to newer VAA datasets to increase the validity of results. Since an election cycle was finished in 2025, more data should be available. We would also like to ablate the effect of English system prompts, that could increase performance, as well as a further exploration of the prompt space.

10. Limitations

This study did not do selected experiments on prompting. As such, better prompting could give more accurate results. Especially when it comes to asking models to be neutral or objective, such prompting could goad the models to answer in a centrist way. Furthermore, we did not attempt to use English system prompts, which has been found to yield better performance also for purely Norwegian tasks.

11. Ethics Statement

We used a dataset for 2021 and published the findings after the 2025 election in order not to influence the election cycle.

12. Acknowledgments

We are grateful to Lemei Zhang for helpful feedback on the manuscript. We would also like to thank reviewers for helpful comments and pointers to future work.

13. Bibliographical References

Mats Faulborn, Indira Sen, Max Pellert, Andreas Spitz, and David Garcia. 2025. Only a little to the left: A theory-grounded measure of political bias in large language models. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 31684–31704, Vienna, Austria. Association for Computational Linguistics.

Shangbin Feng, Chan Young Park, Yuhan Liu, and Yulia Tsvetkov. 2023. 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)*, pages 11737–11762, Toronto, Canada. Association for Computational Linguistics.

Isabel O. Gallegos, Ryan A. Rossi, Joe Barrow, Md Mehrab Tanjim, Sungchul Kim, Franck Dernoncourt, Tong Yu, Ruiyi Zhang, and Nesreen K. Ahmed. 2024. Bias and fairness in large language models: A survey. *Computational Linguistics*, 50(3):1097–1179.

Diego Garzia and Stefan Marschall, editors. 2014. *Matching Voters with Parties and Candidates: Voting Advice Applications in a Comparative Perspective*. ECPR Press, Colchester.

Jochen Hartmann, Jasper Schwenzow, and Maximilian Witte. 2023. The political ideology of conversational ai: Converging evidence on chatgpt's pro-environmental, left-libertarian orientation.

Ruibo Liu, Chenyan Jia, Jason Wei, Guangxuan Xu, and Soroush Vosoughi. 2022. Quantifying and alleviating political bias in language models. *Artificial Intelligence*, 304:103654.

Riccardo Lunardi, David La Barbera, and Kevin Roitero. 2024. The elusiveness of detecting political bias in language models. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24)*. ACM.

Elisabeth Helene Omarhaug. 2016. Valgomater: informasjon eller underholdning? en kvalitativ og kvantitativ studie av valgomatenes funksjoner. Master's thesis, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. Available at: `https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2421155`.

Luca Rettenberger, Markus Reischl, and Mark Schutera. 2025. Assessing political bias in large language models. *Journal of Computational Social Science*, 8:42. Published online: 28 February 2025.

Julian Rutinowski et al. 2025. A cautionary tale about “neutrally” informative ai tools ahead of the 2025 federal elections in germany. In *Explainable Artificial Intelligence - Third World Conference, xAI 2025, Istanbul, Turkey, July 9-11, 2025, Proceedings, Part V*, pages 64–85. Springer. Authors include Sebastian Rieger in related listings; confirm full author list via DOI.

Winnie Street, John Oliver Siy, Geoff Keeling, Adrien Baranes, Benjamin Barnett, Michael McKibben, Tatenda Kanyere, Alison Lentz, Blaise Agüera y Arcas, and Robin I. M. Dunbar. 2026. Llms achieve adult human performance on higher-order theory of mind tasks. *Frontiers in Human Neuroscience*, Volume 19 - 2025.