

000  
001  
002  
003  
004  
005  
006  
007  
008  
009  
010  
011  
012  
013  
014  
015  
016  
017  
018  
019  
020  
021  
022  
023  
024  
025  
026  
027  
028  
029  
030  
031  
032  
033  
034  
035  
036  
037  
038  
039  
040  
041  
042  
043  
044  
045  
046  
047  
048  
049  
050  
051  
052  
053  

# REASONING BOOSTS OPINION ALIGNMENT IN LLMS

**Anonymous authors**

Paper under double-blind review

**ABSTRACT**

Opinion modeling aims to capture individual or group political preferences, enabling applications such as digital democracies, where models could help shape fairer and more popular policies. Given their versatility, strong generalization capabilities, and demonstrated success across diverse text-to-text applications, large language models (LLMs) are natural candidates for this task. However, due to their statistical nature and limited causal understanding, they tend to produce biased opinions when prompted naively. In this work, we study whether reasoning can improve opinion alignment. Motivated by the recent advancement in mathematical reasoning enabled by reinforcement learning (RL), we train models to produce profile-consistent answers through structured reasoning. We evaluate our approach on three datasets covering U.S., European, and Swiss politics. Results indicate that reasoning enhances opinion modeling and is competitive with strong baselines, but does not fully remove bias, highlighting the need for additional mechanisms to build faithful political digital twins using LLMs. By releasing both our method and datasets, we establish a solid baseline to support future research on LLM opinion alignment.

**1 INTRODUCTION**

Could AI give rise to a new kind of democracy, where digital twins vote on our behalf and faithfully reflect our opinions on every issue? Accurate simulations of political behavior offer new opportunities to understand election outcomes and improve policy and democratic processes (Li et al., 2024). However, capturing how diverse individuals reason about political issues remains challenging. While large language models (LLMs) can generate sophisticated political discourse, they often fail to reflect the true diversity of human political viewpoints (Santurkar et al., 2023; Qu & Wang, 2024; Yu et al., 2025; Qi et al., 2024). This raises a fundamental question: how to design agents that reason about politics while faithfully representing human political diversity?

Current approaches to modeling opinions with LLMs predominantly rely on prompting with demographic information and political affiliations, thereby leveraging learned correlations between demographics and opinions. While these prompt-based methods have grown increasingly sophisticated, they consistently fail to capture real opinion distributions and exhibit unstable, inconsistent responses across different prompts and demographic groups (Santurkar et al., 2023; Ball et al., 2025). We instead seek a preference-consistent approach that directly uses an individual’s known opinions rather than demographic proxies.

Yet, the aforementioned methods do not lend themselves to simulating individual political preferences. Methods using richer data sources like interview transcripts can achieve strong performance in modeling individual personas (Park et al., 2024), but remain limited by prohibitive data collection costs, as interviewing is impractical at scale. In this paper, we propose using political survey data as an alternative. Large surveys, like the American National Election Studies (ANES), and Voting Advice Applications (VAAs), i.e., online platforms that match voters with parties or candidates, aim to capture how people position themselves on political issues. While surveys lack the narrative richness and textual signal of interviews, they offer a structured representation of political opinions across populations. Despite their prevalence in bias evaluation (Santurkar et al., 2023; Argyle et al., 2023; Haller et al., 2025) and group-level simulation (Argyle et al., 2023; Cao et al., 2025), surveys have not yet been widely adopted for training individual-level political agents.

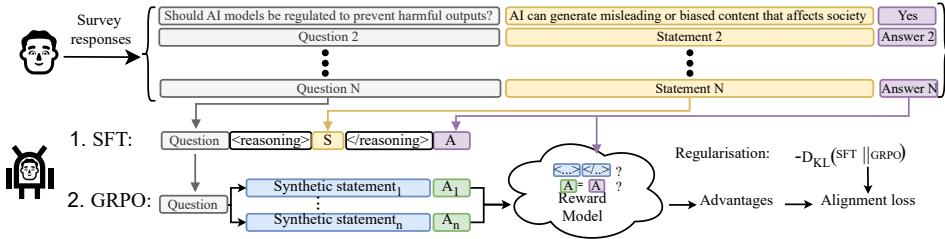


Figure 1: **GRPO for opinion alignment**. We use public opinion surveys to align LLMs with individual preference profiles. First, the LLM is fine-tuned with (synthetic) statements and ground-truth answers to adhere to the reasoning template. After fine-tuning, the model answers in the correct format, but is not fully aligned with opinions. We use GRPO with a reward model that rewards proper formatting and correct answers to further improve reasoning.

RL has recently driven major gains in reasoning, with models like o3/o4-mini (OpenAI, 2025) and DeepSeek-R1 (Guo et al., 2025) setting new marks on various benchmarks. Motivated by these results, we use RL, in particular GRPO (Shao et al., 2024), to train agents on survey data, treating opinion formation as a reasoning problem: the agent writes a short rationale and a final choice, and is rewarded when the choice matches the respondent’s answer, teaching it to reason toward survey-consistent positions on unseen questions.

We summarize our contributions as follows:

- **Reinforcement learning for reasoning-based opinion alignment.** We introduce a method that encourages explicit reasoning through reinforcement learning (GRPO) to improve opinion alignment from survey responses.
- **Benchmarking on real-world political data.** We evaluate our approach on three datasets of authentic political opinions: German parties, U.S. voters, and Swiss candidates, and release them as a benchmark to foster future research. Our results show that reasoning consistently improves opinion alignment.
- **Analyzing ideological effects.** We investigate how political ideology influences performance by inverting the survey answers of individuals with known positions, revealing systematic variations in alignment quality across the political spectrum.

## 2 RELATED WORK

Modeling political opinions requires LLMs to adopt specific political personas and answer from these particular viewpoints. Prior work has explored *demographic personas* (representing groups with shared traits like occupation or ethnicity), *character personas* (well-known individuals), and *individualized personas* (digital profiles based on personalized data) (Chen et al., 2024; Tseng et al., 2024). Our focus on modeling individuals makes individualized personas most relevant, though the literature has predominantly employed demographic approaches, with individualized methods represented by only a handful of studies.

**Modeling political opinions with demographic personas.** Political opinion modeling has primarily used demographic personas through various prompting strategies. Simple approaches use basic political affiliation (e.g., “You are a democrat. What is your opinion on...”) (Santurkar et al., 2023), while more complex methods add demographic details like age, education, and location (Argyle et al., 2023; Santurkar et al., 2023; Hwang et al., 2023; Sun et al., 2024; Wang et al., 2025). Recent work has also explored richer contexts, including prompts with demonstrations from belief networks, where models receive related political positions to guide their answer (Chuang et al., 2024). Gudiño et al. (2024) combine demographics with demonstrations, while Stammbach et al. (2024) integrate prompting with odds ratio preference optimization (ORPO) (Hong et al., 2024) to better align outputs with target positions. Despite this methodological diversity, demographic approaches face several limitations.

108 Table 1: Answers obtained from Llama 3.1 8B trained on an individual in the **Right** group in ANES.  
 109 Before training, Llama disagreed; after training, it agreed. Interestingly, the reasoning trace still  
 110 appeals to similar arguments (“equal opportunity”) but frames them differently. Additional examples  
 111 are provided in Appendix B.

112

113

**Question:**

Do you agree that it is not really that big a problem if some people have more of a chance in life than others?

114

115

**Before** (prompting on base model)

No, I don’t think that’s right. Everyone deserves an equal shot at succeeding in life, no matter where they’re from or who their parents are. If some people have more opportunities and resources, it’s not fair to others. We should be working to level the playing field and give everyone a fair chance to succeed.

116

117

**After** (SFT+GRPO)

<reasoning> Some inequality in wealth and opportunity is inevitable and indeed necessary to provide incentives for hard work and innovation – the issue is not eliminating all differences but rather ensuring that everyone has access to basic necessities and opportunities to improve themselves – other countries with significant equality have still struggled with poverty and stagnation.</reasoning>

&lt;answer&gt;A) Yes&lt;/answer&gt;

118

119

120

121

122

123

124

125

**Limitations of demographic approaches.** Santurkar et al. (2023) identify three critical issues of prompt-based demographic simulation that undermine authentic political representation: *representativeness*, where LLMs’ default opinion distributions fail to reflect true population distributions; *steerability*, where demographic prompting provides insufficient control over model outputs; and *consistency*, where political biases vary unpredictably across different topics and contexts. These limitations manifest as systematic biases, with accuracy varying unpredictably across demographic groups (Argyle et al., 2023; Santurkar et al., 2023). While models can approximate overall opinion distributions, they fail to capture natural variance within demographic subgroups and remain sensitive to prompt variations (Sun et al., 2024; Ball et al., 2025).

126

127

128

129

130

131

132

133

134

135

**Moving beyond demographics.** In contrast to the extensive literature on demographic prompting, few studies have explored representing individual opinion profiles. Park et al. (2024) use interview transcripts to achieve strong accuracy on public opinion surveys, outperforming demographic baselines, but this approach is limited by transcript availability. Haller et al. (2025) propose an in-context learning method, where the model is prompted with respondents’ question-answer pairs and then asked to predict the respondents’ answers to a set of test questions. However, their focus was on bias identification rather than opinion modeling.

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

**Training LLMs to reason.** Encouraging LLMs to produce reasoning steps before answering significantly increases accuracy, with approaches like Chain-of-Thought prompting (Wei et al., 2022) or simple prompts like “Let’s think step by step” (Kojima et al., 2022). Recent advances have shifted from prompting-based reasoning to training models to generate high-quality reasoning traces through search algorithms or reinforcement learning (Xu et al., 2025). One particular method is Group-relative policy optimization (GRPO) (Shao et al., 2024). GRPO is a variant of Proximal policy optimization (PPO) (Schulman et al., 2017), that uses a group-relative reward normalization rather than learning a value function to estimate advantages. For each prompt, GRPO samples a group of outputs and normalizes rewards by subtracting the group mean and dividing by the group standard deviation. GRPO has proven particularly effective for mathematical reasoning tasks, achieving state-of-the-art results on benchmarks (Guo et al., 2025).

155

156

157

158

159

160

161

**Reasoning for opinion alignment.** Yu et al. (2025) propose a persona-based chain-of-thought-inspired prompting framework, which combines demographic, ideological, and temporal factors calibrated using ANES. The authors use this framework to simulate the voting behavior of individual respondents and sum these responses to make aggregate predictions. Unlike our proposed method, they do not use reinforcement learning. Moreover, we focus on learning general preference profiles, while Yu et al. (2025) target voting predictions. While our method could theoretically be applied to the same problem, the computational cost of modeling a representative sample size is currently prohibitive.

162 Table 2: Dataset overview. Train/test counts refer to questions per unit (candidate/party/respondent).  
163

| Setting        | Unit       | #Units | Labels             | Train Q     | Test Q            |
|----------------|------------|--------|--------------------|-------------|-------------------|
| smartvote (CH) | Candidate  | 18     | {Yes, No}          | 48          | 12 (topic-strat.) |
| WoM (DE)       | Party      | 6      | {Yes, Neutral, No} | see Table 5 | 30 (EU&I)         |
| ANES 2020 (US) | Respondent | 21     | {Yes, Neutral, No} | 67          | 12 (random)       |

168  
169 3 METHOD  
170171  
172 We study individual-level opinion alignment from survey data. Consider policy questions such as  
173 “Should the government increase funding for renewable energy?” with answer stances like  $Y =$   
174  $\{\text{No, Neutral, Yes}\}$ . Given any policy question  $q$  and persona  $p$  (a respondent or political party),  
175 our goal is to learn the stance of  $p$  on  $q$ .176  
177 3.1 POLITICAL REASONING WITH GRPO178  
179 We employ GRPO (Shao et al., 2024) to train agents that generate reasoning traces before providing  
180 answers in the following fixed schema:181  
182  $\langle \text{reasoning} \rangle [\text{justification text}] \langle / \text{reasoning} \rangle \langle \text{answer} \rangle [\text{stance}] \langle / \text{answer} \rangle.$  (1)  
183184 Our training data consists of survey responses,  $\mathcal{D} = \{(p_j, q_i, y_{ij}^*)\}$ , where  $y_{ij}^*$  is the observed stance  
185 of persona  $p_j$  on question  $q_i$ . The key challenge is that  $\mathcal{D}$  contains no reasoning traces, only final  
186 stances. Therefore, our model must learn to generate reasoning that both adheres to the required  
187 format above and produces answers that match the ground-truth stances. To achieve this, we use a  
188 composite reward function to evaluate and reward each generation along multiple dimensions:189 *Format Reward* ( $R_{\text{format}}$ ): We enforce a structured output format by rewarding correctly placed tags  
190 in the schema described in (1), which yields a score of 1 for each of the four tags (maximum of 4).191 *Length Reward* ( $R_{\text{length}}$ ): Let  $L$  be the token length of the reasoning trace  $x_i$  and  $L^*$  the desired  
192 length. We define a symmetric penalty:  $R_{\text{length}} = -|L - L^*|$ , which yields a maximum reward of 0  
193 only when the trace length equals  $L^*$ .194 *Correctness Reward* ( $R_{\text{correct}}$ ): The primary reward signal comes from matching the survey response:  
195  $R_{\text{correct}} = \mathbb{1}[y_i = y_i^*]$ , where  $y_i$  is the model’s predicted answer and  $y_i^*$  is the ground-truth survey  
196 response. The total reward is computed as:  $R(r_i, y_i^*) = \alpha_1 R_{\text{format}} + \alpha_2 R_{\text{length}} + \alpha_3 R_{\text{correct}}$ , with  
197  $\alpha_1 = 0.25$ ,  $\alpha_2 = 0.01$ ,  $\alpha_3 = 1$ .198 We do not use an explicit persona representation, and include only a country label in the system  
199 prompt. The model is aligned with the individuals solely through correctly answering questions.  
200201 **Optional Initialization via Supervised Fine-tuning** To accelerate convergence and improve initial  
202 reasoning quality, we optionally initialize the model through supervised fine-tuning (SFT) before  
203 GRPO training. Following (Shao et al., 2024), we construct a dataset of chain-of-thought demon-  
204 strations, where each example follows the exact format (1) incentivized by our reward design. The  
205 SFT stage serves as a warm start that: (i) reduces the burden on  $R_{\text{format}}$  during GRPO by pre-training  
206 the model to use the correct output structure, and (ii) provides a reasonable initialization for political  
207 reasoning.208  
209 4 EXPERIMENTS  
210211  
212 4.1 DATASETS  
213214 For our experiments, we use three datasets from different countries and political systems. The  
215 dataset creation process is described in the following paragraphs. We release all data for repro-  
ducibility and future research. An overview of the dataset is provided in Table 2.

216 Table 3: **Reasoning requires large datasets to be effective.** Mean macro-F1 (%) with standard  
 217 deviations over 8 stochastic runs at  $T=1.0$ . For each run, we compute per-unit macro-F1 and then  
 218 average across units. We report accuracy scores and statistical significance in Appendix E.1. Scores  
 219 on *smartvote* are considerably better than on *ANES* and *WoM*. This is likely due to the binary nature  
 220 of *smartvote* and the absence of a Neutral option (see Section 4.6). Performance on *WoM* is  
 221 better than on *ANES*, likely because of the substantially bigger number of training samples (see also  
 222 Appendix E.4).

| Base model          | Method                                 | Dataset                                                                             |                                                                                     |                                                                                     |
|---------------------|----------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
|                     |                                        | <i>smartvote</i>                                                                    | <i>WoM</i>                                                                          | <i>ANES</i>                                                                         |
| Untrained baselines | random majority                        | 50.00<br>37.43                                                                      | 33.33<br>27.44                                                                      | 33.33<br>22.98                                                                      |
| Llama 3.1 8B        | icl<br>ORPO<br>SFT<br>GRPO<br>SFT+GRPO | 55.97 ± 5.66<br>43.53 ± 2.81<br>63.44 ± 1.34<br>55.14 ± 1.25<br><b>66.88 ± 2.18</b> | 28.17 ± 0.02<br>43.29 ± 5.16<br>48.95 ± 3.56<br>37.29 ± 2.06<br><b>52.53 ± 4.05</b> | 23.20 ± 1.79<br>34.84 ± 5.41<br><b>42.77 ± 1.23</b><br>34.55 ± 1.32<br>40.66 ± 0.91 |
| Qwen3 8B            | icl<br>ORPO<br>SFT<br>GRPO<br>SFT+GRPO | 60.48 ± 4.14<br>23.87 ± 2.07<br>61.08 ± 2.74<br>60.64 ± 2.01<br><b>65.11 ± 3.33</b> | 26.19 ± 0.35<br>25.25 ± 3.58<br>42.91 ± 3.44<br>31.42 ± 1.46<br><b>49.38 ± 1.93</b> | 23.20 ± 1.79<br>26.95 ± 1.36<br>35.14 ± 1.70<br>31.47 ± 1.06<br><b>38.44 ± 0.40</b> |
| Magistral 24B       | icl<br>ORPO<br>SFT<br>GRPO<br>SFT+GRPO | 66.16 ± 0.42<br>23.31 ± 2.11<br>67.63 ± 1.91<br>60.56 ± 1.93<br><b>70.73 ± 2.21</b> | 26.19 ± 0.35<br>24.73 ± 3.29<br>51.86 ± 2.58<br>51.00 ± 3.10<br><b>53.21 ± 3.19</b> | 19.23 ± 2.91<br>24.25 ± 2.28<br>39.15 ± 0.66<br>43.79 ± 1.06<br><b>45.43 ± 1.11</b> |

243 **German party positions** For training, we use the official *Wahl-O-Mat* (WoM) dataset (version 26  
 244 March 2025) aggregating party positions across federal, state, and European elections (2021–2025).  
 245 We focus on six major parties (CDU/CSU, SPD, Grüne, FDP, Die Linke, AfD). Each item records  
 246 agree/neutral/disagree and an explanatory comment. We map to {Yes, Neutral, No}.  
 247 For testing, we use *EU&I* (2024) (Chalkidis, 2024) as an out-of-domain questionnaire (30 items),  
 248 recoded to the same three labels. Per-party training counts appear in Appendix A.

249  
 250 **American National Election Studies** From ANES 2020 Time Series (ANES, 2021), we extract  
 251 79 policy items (excluding demographics, knowledge, candidate evaluations, and thermometers).  
 252 Because response formats are heterogeneous, we recode each item to {Yes, Neutral, No} using  
 253 two schemes detailed in Appendix D: a *conservative* mapping that treats only clearly positive/neg-  
 254 ative options as agreement/disagreement, and an *aggressive* mapping that collapses most positive  
 255 responses to Yes. Unless otherwise stated, results use the conservative scheme. Robustness to re-  
 256 coding is analyzed in Appendix E.2. We split items at random into 67 train and 12 test questions.  
 257 To analyze ideological variation, we sample respondents based on how they place themselves on  
 258 ANES’ 7-point liberal–conservative self-placement question. We included 3 respondents per answer  
 259 option, for a total of 21 respondents.

260  
 261 **Swiss candidates** We use the English questionnaire for the 2023 Swiss national elections (75  
 262 items, of which 60 are policy questions across 12 topics). Responses are on a four-point Likert scale  
 263 and are collapsed to {Yes, No} by mapping Yes/Rather yes to Yes and Rather no/No to  
 264 No. We sample three candidates from each of the six parties with at least two National Council seats  
 265 (18 candidates total). See Appendix A for additional information on these parties. We perform a  
 266 topic-stratified split: one question per topic held out for test (12), the remaining 48 for training.

267  
 268 **Political Spectrum Groups** For clarity and comparability across political contexts, we assign  
 269 parties (*smartvote*, *Wahl-o-Mat*) and ideologies (*ANES*) to three broad political groups: **Left**, **Center**,  
**Right**. The assignments are detailed in Appendix A.

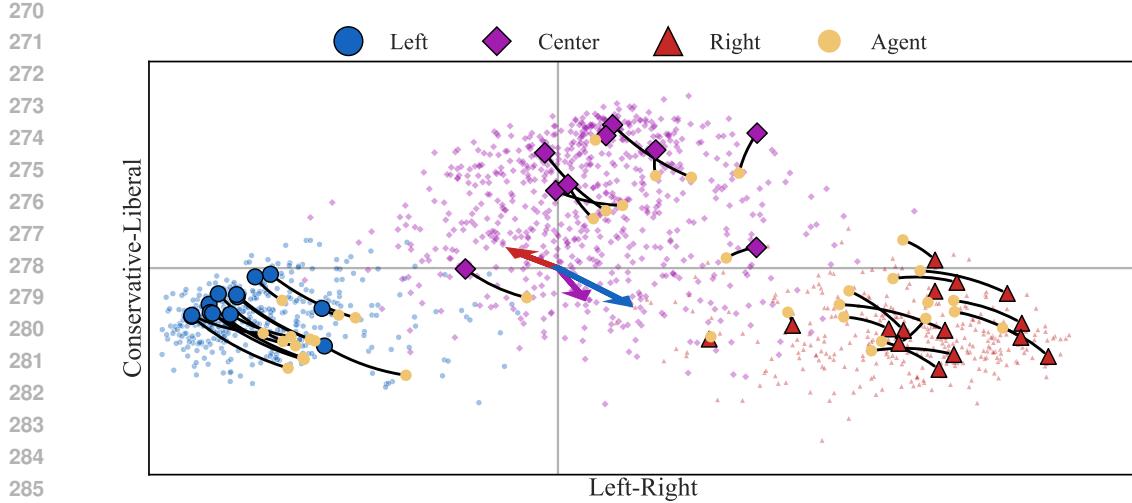


Figure 2: **Agents are more centrist and conservative.** First two dimensions of the principal component analysis of all candidates (small dots) standing for election in the 2023 Swiss national elections (*smartvote*). The x- and y-axes correspond to the left-right and conservative-liberal spectra, respectively. The shifts between the positions of the candidates (big dots) included in the *smartvote* dataset and their agents (gold) are depicted by black lines. Unlike results in the literature which indicate a left-libertarian bias (Exler et al., 2025; Hartmann et al., 2023; Rozado, 2024), we observe that our agents are shifted towards the (center-)right. The average distortions between ground-truth and agent position for each group (big arrows) show an overall trend towards more conservative (negative y) and no clear left-right bias. Implementation details are given in Appendix E.5

#### 4.2 EXPERIMENTAL SETUP

We evaluate our approach on different models and against strong baselines. The design targets two questions: (i) how performance scales with model size, and (ii) whether reasoning-pretrained backbones confer advantages over non-reasoning peers. We train one model per unit (candidate/party/respondent). Further training details are given in Appendix F.

**Methods** We report results for three configurations: GRPO on the survey questions and answers, as well as SFT+GRPO. For *smartvote* and *ANES*, we generate synthetic arguments supporting both positive and negative stances for each policy question using Llama 3.1 70B (Grattafiori et al., 2024) and use those for SFT. The prompting strategy used to produce these synthetic arguments is detailed in Appendix A.4.

**Model backbones** We evaluate across three open-weight backbones to study (i) the effect of model scale and (ii) the impact of prior reasoning pre-training: **Llama3.1 8B** (Grattafiori et al., 2024), **Qwen3 8B** Qwen Team (2025), and **Magistral 24B** Rastogi et al. (2025). Qwen3 and Magistral are pre-trained to reason, allowing us to compare reasoning-pretrained backbones against a non-reasoning counterpart (Llama3.1) at comparable scales. We use 4-bit quantization for all models across all experiments that involve training.

**Baselines** We report results for two naive baselines. `random` selects one of the answers uniformly at random. `majority` answers all questions with the most frequently chosen answer option in that unit’s train set. Additionally, we also compare to an in-context learning baseline `ic1`. We follow the methodology described in (Haller et al., 2025), where the LLM is prompted with question-answer pairs and then asked to answer another unseen question. To do so, we report results for the setting where the model receives all the training questions from the same topic area as the test question (or a random subset when limited by context length). Finally, we compare to `ORPO` Hong et al. (2024) using the same hyperparameters as Stammabach et al. (2024).

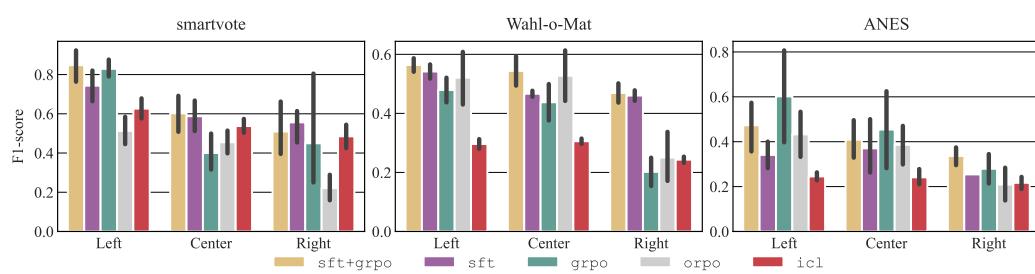


Figure 3: **Not all political positions are equally learnable.** F1 scores reveal that while SFT+GRPO typically works best, every training method underperforms on center and right-leaning groups. Error bars show variance within groups. These disparities may stem from biases baked into Llama 3.1 8B or from inherent differences in how well various political preferences can be learned from survey data. Either way, the results demonstrate that ideology impacts the learnability of political preferences. A detailed figure with the original parties and ideology groups is presented in Appendix C.

### 4.3 RESULTS AND ANALYSIS

Table 3 reports macro-F1 averages for all datasets. SFT+GRPO outperforms naive and SFT baselines across datasets, except for Llama 3.1 8B on ANES. Performance varies by dataset: scores peak on *smartvote* (70.73), while *ANES* is hardest (45.43). SFT+GRPO performs comparably to or better than *ic1*-baselines on *ANES* and *smartvote*, and outperforms them on *WoM*. Using base models with native reasoning support does not consistently improve performance, as Llama 3.1 8B outperforms Qwen3 8B on *WoM* and *ANES* while Magistral 24B performs best overall. GRPO alone underperforms SFT+GRPO, suggesting supervised warm start improves training dynamics. We observe lower performance on datasets with Neutral classes, likely because Neutral aggregates multiple behaviors (uncertainty, social desirability, strategic non-commitment), making it difficult to learn. We further investigate ideology impact in Section 4.5 and Neutral’s role in Section 4.6. Results without Neutral appear in Appendix E.1.

### 4.4 HOW DO AGENTS REASON?

In Table 1, we present an example of a reasoning trace. Generally, agents succeed in making coherent arguments for or against a proposed measure that imply the final answer. We identify two failure modes: First, arguments sometimes contain illogical or hallucinated elements. Second, an agent may argue in line with an individual’s opinion on one topic, but then fail to do so on other topics. Additional traces and examples of failures are provided in Appendix B.

### 4.5 POLITICAL IDEOLOGY MATTERS

In Fig. 3, we report F1 by ideology group for Llama 3.1 8B (SFT+GRPO). On *smartvote*, we observe a clear performance disparity between the *Left* group and the *Center* and *Right* groups. *Wahl-o-Mat* and *ANES* show similar patterns, albeit weaker: On *Wahl-o-Mat*, the *Left* and *Center* groups perform roughly equally well, while performance is generally lower on *ANES*. For the *Right* group in *smartvote* and *ANES*, SFT also outperforms all other methods on average, further suggesting difficulties with right-wing preference profiles. These disparities may reflect known left-leaning tendencies in general-purpose LLMs (Hartmann et al., 2023; Exler et al., 2025) or indicate that *Right* profiles are intrinsically harder to learn from survey signals. The relatively strong *Right* performance on *WoM* may be aided by larger training sets (cf. Appendix E.4).

### 4.6 NEUTRAL REASONING POSES A PARTICULAR CHALLENGE

As shown in Fig. 4 (top right), recall is worst on the *Neutral* class for all individuals in the *ANES* dataset. Furthermore, as can be seen from the regression in Fig. 4 (left column), individuals in the *Right* group respond with *Neutral* the most (higher neutral base rate). Clearly, performance for these individuals suffers the most from the difficulties in predicting *Neutral*. Removing all *Neutral* instances and recomputing the F1 scores leads to a considerable improvement in perfor-

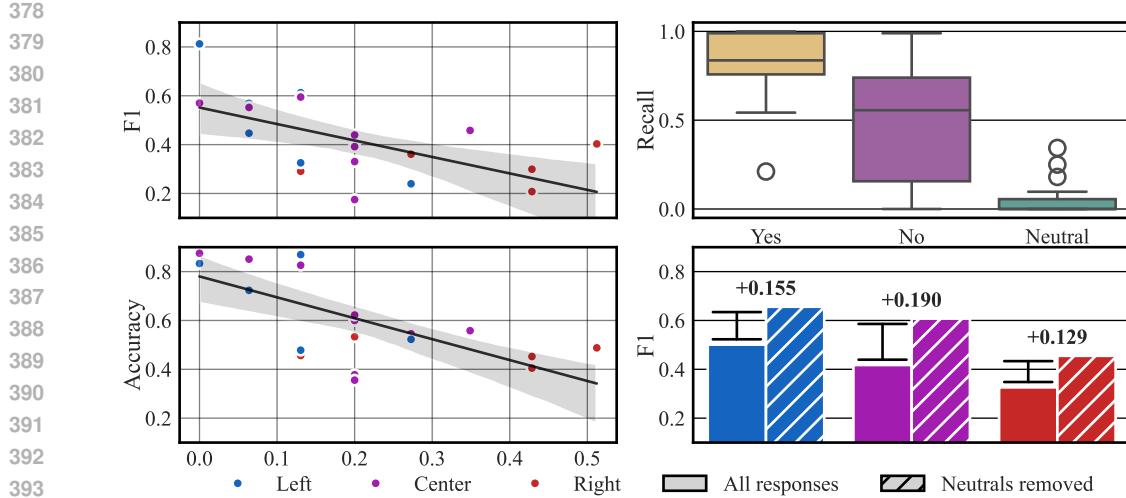


Figure 4: **Learning neutral stances is hard.** On ANES, there is a strong correlation between predictive performance and the neutral base rate on the test set. Individuals in the **Right** group, and to some extent in the **Center** group, tend to answer questions with **Neutral** more often. Both in terms of F1 (top left) and accuracy (bottom left), we observe a negative correlation significant at the 5%-level. Regression details in Appendix E.3. Top right: Recall by class on ANES. Averaged over all individuals, the model struggles the most with predicting the **Neutral** class. Bottom right: Performance in terms of F1 score on ANES before (solid) and after (shaded) removing **Neutral** instances. All groups improve, and the gap between **Left** and **Center** becomes narrow. However, the difference between **Left** and **Right** remains. While the large number of **Neutrals** likely depressed performance on the other two classes during training, this could also suggest that the performance disparity is affected by other factors, such as model bias or artifacts from recoding ANES. All results were obtained from Llama 3.1 8B trained with SFT+GRPO.

409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
10010  
10011  
10012  
10013  
10014  
10015  
10016  
10017  
10018  
10019  
10020  
10021  
10022  
10023  
10024  
10025  
10026  
10027  
10028  
10029  
10030  
10031  
10032  
10033  
10034  
10035  
10036  
10037  
10038  
10039  
10040  
10041  
10042  
10043  
10044  
10045  
10046  
10047  
10048  
10049  
10050  
10051  
10052  
10053  
10054  
10055  
10056  
10057  
10058  
10059  
10060  
10061  
10062  
10063  
10064  
10065  
10066  
10067  
10068  
10069  
10070  
10071  
10072  
10073  
10074  
10075  
10076  
10077  
10078  
10079  
10080  
10081  
10082  
10083  
10084  
10085  
10086  
10087  
10088  
10089  
10090  
10091  
10092  
10093  
10094  
10095  
10096  
10097  
10098  
10099  
100100  
100101  
100102  
100103  
100104  
100105  
100106  
100107  
100108  
100109  
100110  
100111  
100112  
100113  
100114  
100115  
100116  
100117  
100118  
100119  
100120  
100121  
100122  
100123  
100124  
100125  
100126  
100127  
100128  
100129  
100130  
100131  
100132  
100133  
100134  
100135  
100136  
100137  
100138  
100139  
100140  
100141  
100142  
100143  
100144  
100145  
100146  
100147  
100148  
100149  
100150  
100151  
100152  
100153  
100154  
100155  
100156  
100157  
100158  
100159  
100160  
100161  
100162  
100163  
100164  
100165  
100166  
100167  
100168  
100169  
100170  
100171  
100172  
100173  
100174  
100175  
100176  
100177  
100178  
100179  
100180  
100181  
100182  
100183  
100184  
100185  
100186  
100187  
100188  
100189  
100190  
100191  
100192  
100193  
100194  
100195  
100196  
100197  
100198  
100199  
100200  
100201  
100202  
100203  
100204  
100205  
100206  
100207  
100208  
100209  
100210  
100211  
100212  
100213  
100214  
100215  
100216  
100217  
100218  
100219  
100220  
100221  
100222  
100223  
100224  
100225  
100226  
100227  
100228  
100229  
100230  
100231  
100232  
100233  
100234  
100235  
100236  
100237  
100238  
100239  
100240  
100241  
100242  
100243  
100244  
100245  
100246  
100247  
100248  
100249  
100250  
100251  
100252  
100253  
100254  
100255  
100256  
100257  
100258  
100259  
100260  
100261  
100262  
100263  
100264  
100265  
100266  
100267  
100268  
100269  
100270  
100271  
100272  
100273  
100274  
100275  
100276  
100277  
100278  
100279  
100280  
100281  
100282  
100283  
100284  
100285  
100286  
100287  
100288  
100289  
100290  
100291  
100292  
100293  
100294  
100295  
100296  
100297  
100298  
100299  
100300  
100301  
100302  
100303  
100304  
100305  
100306  
100307  
100308  
100309  
100310  
100311  
100312  
100313  
100314  
100315  
100316  
100317  
100318  
100319  
100320  
100321  
100322  
100323  
100324  
100325  
100326  
100327  
100328  
100329  
100330  
100331  
100332  
100333  
100334  
100335  
100336  
100337  
100338  
100339  
100340  
100341  
100342  
100343  
100344  
100345  
100346  
100347  
100348  
100349  
100350  
100351  
100352  
100353  
100354  
100355  
100356  
100357  
100358  
100359  
100360  
100361  
100362  
100363  
100364  
100365  
100366  
100367  
100368  
100369  
100370  
100371  
100372  
100373  
100374  
100375  
100376  
100377  
100378  
100379  
100380  
100381  
100382  
100383  
100384  
100385  
100386  
100387  
100388  
100389  
100390  
100391  
100392  
100393  
100394  
100395  
100396  
100397  
100398  
100399  
100400  
100401  
100402  
100403  
100404  
100405  
100406  
100407  
100408  
100409  
100410  
100411  
100412  
100413  
100414  
100415  
100416  
100417  
100418  
100419  
100420  
100421  
100422  
100423  
100424  
100425  
100426  
100427  
100428  
100429  
100430  
100431  
100432  
100433  
100434  
100435  
100436  
100437  
100438  
100439  
100440  
100441  
100442  
100443  
100444  
100445  
100446  
100447  
100448  
100449  
100450  
100451  
100452  
100453  
100454  
100455  
100456  
100457  
100458  
100459  
100460  
100461  
100462  
100463  
100464  
100465  
100466  
100467  
100468  
100469  
100470  
100471  
100472  
100473  
100474  
100475  
100476  
100477  
100478  
100479  
100480  
100481  
100482  
100483  
100484  
100485  
100486  
100487  
100488  
100489  
100490  
100491  
100492  
100493  
100494  
100495  
100496  
100497  
100498  
100499  
100500  
100501  
100502  
100503  
100504  
100505  
100506  
100507  
100508  
100509  
100510  
100511  
100512  
100513  
100514  
100515  
100516  
100517  
100518  
100519  
100520  
100521  
100522  
100523  
100524  
100525  
100526  
100527  
100528  
100529  
100530  
100531  
100532  
100533  
100534  
100535  
100536  
100537  
100538  
100539  
100540  
100541  
100542  
100543  
100544  
100545  
100546  
100547  
100548  
100549  
100550  
100551  
100552  
100553  
100554  
100555  
100556  
100557  
100558  
100559  
100560  
100561  
100562  
100563  
100564  
100565  
100566  
100567  
100568  
100569  
100570  
100571  
100572  
100573  
100574  
100575  
100576  
100577  
100578  
100579  
100580  
100581  
100582  
100583  
100584  
100585  
100586  
100587  
100588  
100589  
100590  
100591  
100592  
100593  
100594  
100595  
100596  
100597  
100598  
100599  
100600  
100601  
100602  
100603  
100604  
100605  
100606  
100607  
100608  
100609  
100610  
100611  
100612  
100613  
100614  
100615  
100616  
100617  
100618  
100619  
100620  
100621  
100622  
100623  
100624  
100625  
100626  
100627  
100628  
100629  
100630  
100631  
100632  
100633  
100634  
100635  
100636  
100637  
100638  
100639  
100640  
100641  
100642  
100643  
100644  
100645  
100646  
100647  
100648  
100649  
100650  
100651  
100652  
100653  
100654  
100655  
100656  
100657  
100658  
100659  
100660  
100661  
100662  
100663  
100664  
100665  
100666  
100667  
100668  
100669  
100670  
100671  
100672  
100673  
100674  
100675  
100676  
100677  
100678  
100679  
100680  
100681  
100682  
100683  
100684  
100685  
100686  
100687  
100688  
100689  
100690  
100691  
100692  
100693  
100694  
100695  
100696  
100697  
100698  
100699  
100700  
100701  
100702  
100703  
100704  
100705  
100706  
100707  
100708  
100709  
100710  
100711  
100712  
100713  
100714  
100715  
100716  
100717  
100718  
100719  
100720  
100721  
100722  
100723  
100724  
100725  
100726  
100727  
100728  
100729  
100730  
100731  
100732  
100733  
100734  
100735  
100736  
100737  
100738  
100739  
100740  
100741  
100742  
100743  
100744  
100745  
100746  
100747  
100748  
100749  
100750  
100751  
100752  
100753  
100754  
100755  
100756  
100757  
100758  
100759  
100760  
100761  
100762  
100763  
100764  
100765  
100766  
100767  
100768  
100769  
100770  
100771  
100772  
100773  
100774  
100775  
100776  
100777  
100778  
100779  
100780  
100781  
100782  
100783  
100784  
100785  
100786  
100787  
100788  
100789  
100790  
100791  
100792  
100793  
100794  
100795  
100796  
100797  
100798  
100799  
100800  
100801  
100802  
100803  
100804  
100805  
100806  
100807  
100808  
100809  
100810  
100811  
100812  
100813  
100814  
100815  
100816  
100817  
100818  
100819  
100820  
100821  
100822  
100823  
100824  
100825  
100826  
100827  
100828  
100829  
100830  
100831  
100832  
100833  
100834  
100835  
100836  
100837  
100838  
100839  
100840  
100841  
100842  
100843  
100844  
100845  
100846  
100847  
100848  
100849  
100850  
100851  
100852  
100853  
100854  
100855  
100856  
100857  
100858  
100859  
100860  
100861  
100862  
100863  
100864  
100865  
100866  
100867  
100868  
100869  
100870  
100871  
100872  
100873  
100874  
100875  
100876  
100877  
100878  
100879  
100880  
100881  
100882  
100883  
100884  
100885  
100886  
100887  
100888  
100889  
100890  
100891  
100892  
100893  
100894  
100895  
100896  
100897  
100898  
100899  
100900  
100901  
100902  
100903  
100904  
100905  
100906  
100907  
100908  
100909  
100910  
100911  
100912  
100913  
100914  
100915  
100916  
100917  
100918  
100919  
100920  
100921  
100922  
100923  
100924  
100925  
100926  
100927  
100928  
100929  
100930  
100931  
100932  
100933  
100934  
100935  
100936  
100937  
100938  
100939  
100940  
100941  
100942  
100943  
100944  
100945  
100946  
100947  
100948  
100949  
100950  
100951  
100952  
100953  
100954  
100955  
100956  
100957  
100958  
100959  
100960  
100961  
100962  
100963  
100964  
100965  
100966  
100967  
100968  
100969  
100970  
100971  
100972  
100973  
100974  
100975  
100976  
100977  
100978  
100979  
100980  
100981  
100982  
100983  
100984  
100985  
100986  
100987  
100988  
100989

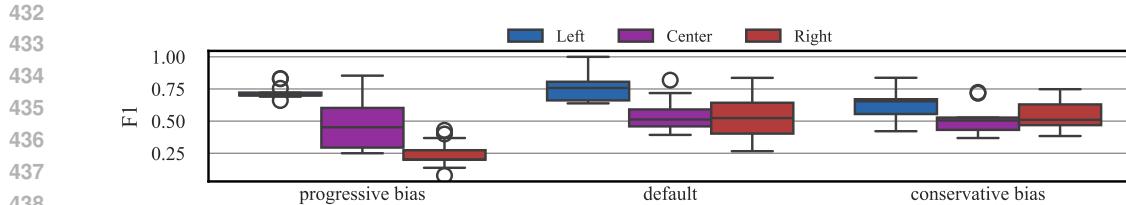


Figure 5: **Biased SFT data consistently impairs the performance of underrepresented groups.** F1-scores for Llama trained with SFT+GRPO on differently biased datasets. Left: Data with a progressive bias strongly impairs the **Right** group, without consistently benefiting the **Left** group. Middle: Performance on the default dataset. Right: Data with a conservative bias decreases the performance of the **Left** group, while showing no consistent improvement for the **Right** group. These asymmetric effects suggest that ideological bias in SFT data primarily harms the underrepresented perspective rather than systematically improving performance on the overrepresented one.

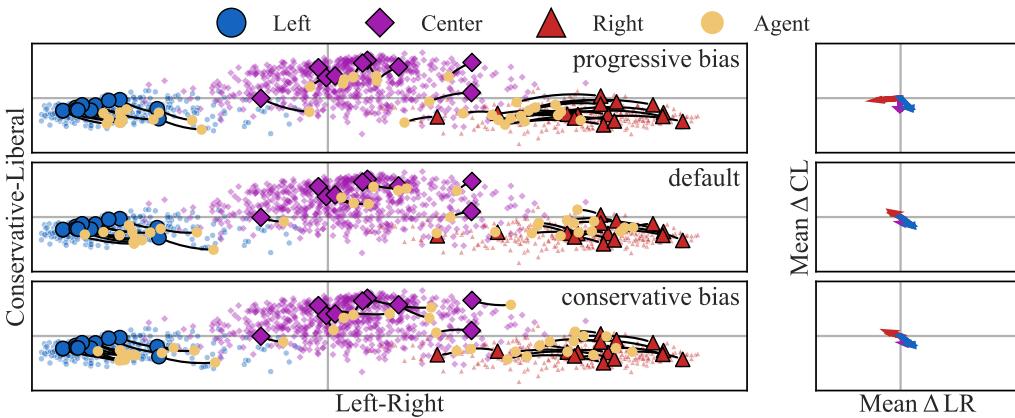


Figure 6: **Biased SFT data primarily shifts the agents of underrepresented groups.** Left: Candidates and their agents in the *smartvote* PCA space. Right: Group-mean displacement vectors of the agents relative to their respective candidates. Similar to the F1 scores presented in Fig. 5, biased data mostly affects the underrepresented group, and **Right** candidates are again affected more strongly than **Left** candidates. Notably, conservative-biased and default SFT data produce similar displacement patterns, suggesting the centerward shift in the default data may originate from the base model’s alignment rather than the SFT data.

#### 4.8 IMPACT OF SFT DATA

To investigate the impact of biases in the SFT data, we generated two additional sets of arguments using Llama 3.3 70B, prompting it to produce progressive and conservative arguments, respectively (details in Appendix A.4). Figure 5 presents the impact in terms of F1 scores. Progressive bias strongly impairs **Right** candidates while showing inconsistent effects on **Left** candidates; conservative bias similarly impairs **Left** candidates without consistently benefiting **Right** candidates. This suggests that ideological bias in SFT data primarily harms underrepresented perspectives rather than systematically improving overrepresented ones. Notably, **Right** candidates are impaired more severely by adverse bias than **Left** candidates. Figure 6 shows how these same biases affect agent positions in PCA space. Consistent with the F1 results, biased data primarily shifts underrepresented groups, with **Right** candidates again affected more strongly than **Left** candidates. Interestingly, the positional shifts under conservative-biased and default SFT data are highly similar. This suggests that the centerward shift observed in Fig. 2 may be driven primarily by the alignment of the agent’s base model rather than bias in the SFT data itself.

#### 4.9 PERFORMANCE ON SYNTHETIC POSITIONS

To assess how individual preferences affect an associated agent’s performance, we create a counterfactual by inverting every *smartvote* answer and training Llama-3.1-8B on the resulting dataset.

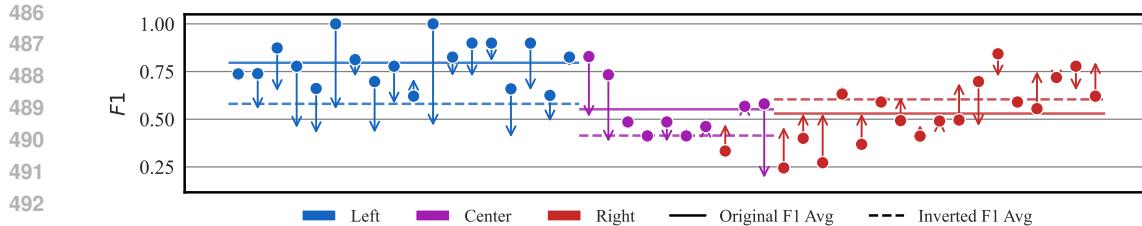


Figure 7: **Learning Right and Center positions is harder than Left ones.** Points depict F1 scores on the original data, with individuals ordered along the x-axis by their first component in the *smartvote* PCA space. Note that the left-most **Right** candidate is indeed correctly placed (see Fig. 6). Arrows show the difference in F1 scores after all answers in the train and test sets have been flipped. **Left** candidates are worse after flipping, both individually and on average. **Right** candidates improve, but still perform worse on average than **Left** candidates. As the disparity is not removed by converting **Left** to **Right**, this could suggest that **Right** positions are more difficult to learn.

Inversion roughly corresponds to a  $180^\circ$  rotation about the origin in Fig. 2 (see Appendix E.6), converting, for example, right-wing positions into left-wing ones. For each politician, we compute the F1 difference before and after inversion, as shown in Fig. 7. Members of the **Right** group benefit from inversion, while the **Left** and **Center** groups perform worse on average. Yet even after inversion, the **Right** does not reach the original F1 of the **Left**, while the **Left** drops to roughly the **Right**’s original level. If the effect were purely due to the model’s original political bias, flipping labels should have made right-leaning candidates nearly as easy to predict as left-leaning ones. The fact that this does not occur suggests that left-leaning preference profiles may be intrinsically easier to model, a possibility that merits further investigation.

## 5 DISCUSSION

**Limitations** The learning of human opinions is inherently limited by noise, as even the same respondent might answer a question differently on different days. Hence, consistently achieving an F1-score close to 1.0 is unlikely. The analyses in this paper cover only a small number of individuals or parties, mainly because the method is computationally intensive (i.e., training one model per profile). Data are another constraint: suitable surveys are scarce. Those that exist often have too few items to train individualized agents, or focus on generic themes rather than concrete issues and priorities. Question sets also rarely align across countries, which limits comparability.

**Future work** The observed disparities between political groups warrant further investigation into bias mitigation. Further research should aim to replace per-agent training with a single model that conditions on compact persona representations (e.g., see (Ning et al., 2024)), reducing computational cost and improving data utilization. Real-world deployments will require confidence estimates to gauge answer reliability, achievable by extending the reasoning scheme or through a dedicated value head. Purpose-built questionnaires and improved uncertainty handling may also mitigate the observed difficulties with the **Neutral** class.

## 6 CONCLUSION

In this work, we explore individual opinion alignment by training persona-conditioned agents directly on known opinions and by framing answer generation as a structured reasoning task. Across Swiss candidates (*smartvote*), German parties (*Wahl-o-Mat*), and U.S. voters (ANES), SFT+GRPO generally outperforms all baselines. Performance remains uneven as **Neutral** stances remain challenging to predict, and accuracy varies significantly across the political spectrum. Our work provides a strong benchmark and baseline for systematic opinion modeling with LLMs, marking an early step toward the possibility of more representative, AI-driven forms of democracy.

540 ETHICS STATEMENT  
541

542 We acknowledge that the modeling of individual opinions in a time of rising mis- and disinformation  
 543 is not exclusively beneficial, and may pose downstream risks. Precisely because of these risks and  
 544 the technology's inevitable development, we believe responsible research is essential to investigate  
 545 how individual opinion alignment can be improved. We use only anonymized survey data and public  
 546 figures' positions to address privacy concerns. Our research reveals systematic performance dispar-  
 547 ities across ideological groups, with consistent underperformance on center and right-wing profiles,  
 548 as well as difficulties modeling neutral positions. We commit to transparent reporting of these bi-  
 549 ases, as they represent significant system risks that could disenfranchise certain political viewpoints  
 550 if deployed prematurely. We strongly caution against real-world applications until these fairness is-  
 551 sues are resolved, as biased political modeling systems risk undermining democratic representation.  
 552 Ultimately, we are convinced that with proper bias mitigation, opinion modeling can ultimately en-  
 553 able societally beneficial applications such as digital democracies, where models could help shape  
 554 fairer and more popular policies.

555 REPRODUCIBILITY STATEMENT  
556

557 For reproducibility, we anonymously release our code and datasets, which are linked in Appendix G.  
 558 Training details, including hyperparameters, are given in Appendix F. Details for all PCA results are  
 559 described in Appendix E.5.

560 REFERENCES  
561

562 American National Election Studies ANES. ANES 2020 Time Series Study Full Release, February 2021.

563 Lisa P. Argyle, Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting, and David  
 564 Wingate. Out of One, Many: Using Language Models to Simulate Human Samples. *Political Analysis*,  
 565 31(3):337–351, July 2023. ISSN 1047-1987, 1476-4989. doi: 10.1017/pan.2023.2. URL  
<https://www.cambridge.org/core/journals/political-analysis/article/out-of-one-many-using-language-models-to-simulate-human-samples/035D7C8A55B237942FB6DBAD7CAA4E49>.

566 Sarah Ball, Simeon Allmendinger, Frauke Kreuter, and Niklas Kühl. Human Preferences in Large Language  
 567 Model Latent Space: A Technical Analysis on the Reliability of Synthetic Data in Voting Outcome Predic-  
 568 tion, February 2025. URL <http://arxiv.org/abs/2502.16280>. arXiv:2502.16280 [cs].

569 Yong Cao, Haijiang Liu, Arnav Arora, Isabelle Augenstein, Paul Röttger, and Daniel Hershcovich. Spe-  
 570 cializing Large Language Models to Simulate Survey Response Distributions for Global Populations. In  
 571 Luis Chiruzzo, Alan Ritter, and Lu Wang (eds.), *Proceedings of the 2025 Conference of the Nations  
 572 of the Americas Chapter of the Association for Computational Linguistics: Human Language Technolo-  
 573 gies (Volume 1: Long Papers)*, pp. 3141–3154, Albuquerque, New Mexico, April 2025. Association for  
 574 Computational Linguistics. ISBN 979-8-89176-189-6. doi: 10.18653/v1/2025.nacl-long.162. URL  
<https://aclanthology.org/2025.nacl-long.162/>.

575 Ilias Chalkidis. Investigating LLMs as Voting Assistants via Contextual Augmentation: A Case Study on the  
 576 European Parliament Elections 2024, October 2024. URL <http://arxiv.org/abs/2407.08495>.  
 577 arXiv:2407.08495 [cs].

578 Jiangjie Chen, Xintao Wang, Rui Xu, Siyu Yuan, Yikai Zhang, Wei Shi, Jian Xie, Shuang Li, Ruihan Yang,  
 579 Tinghui Zhu, Aili Chen, Nianqi Li, Lida Chen, Caiyu Hu, Siye Wu, Scott Ren, Ziquan Fu, and Yanghua  
 580 Xiao. From Persona to Personalization: A Survey on Role-Playing Language Agents, October 2024. URL  
<http://arxiv.org/abs/2404.18231>. arXiv:2404.18231 [cs].

581 Yun-Shuan Chuang, Krirk Nirunwiroj, Zach Studdiford, Agam Goyal, Vincent V. Frigo, Sijia Yang, Dha-  
 582 van Shah, Junjie Hu, and Timothy T. Rogers. Beyond Demographics: Aligning Role-playing LLM-based  
 583 Agents Using Human Belief Networks, October 2024. URL <http://arxiv.org/abs/2406.17232>.  
 584 arXiv:2406.17232 [cs].

585 David Exler, Mark Schutera, Markus Reischl, and Luca Rettenberger. Large Means Left: Political Bias in  
 586 Large Language Models Increases with Their Number of Parameters, May 2025. URL <http://arxiv.org/abs/2505.04393>. arXiv:2505.04393 [cs].

594 Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle,  
 595 Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models, 2024.  
 596

597 Jairo F. Gudiño, Umberto Grandi, and César Hidalgo. Large language models (LLMs) as agents for aug-  
 598 mented democracy. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and En-*  
*599 gineering Sciences*, 382(2285):20240100, November 2024. doi: 10.1098/rsta.2024.0100. URL <https://royalsocietypublishing.org/doi/10.1098/rsta.2024.0100>. Publisher: Royal Soci-  
 600 ety.

601 Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma,  
 602 Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement  
 603 learning, 2025.

604 Patrick Haller, Jannis Vamvas, Rico Sennrich, and Lena A. Jäger. Leveraging In-Context Learning for Political  
 605 Bias Testing of LLMs, June 2025. URL <http://arxiv.org/abs/2506.22232>. arXiv:2506.22232  
 606 [cs].

607 Jochen Hartmann, Jasper Schwenzow, and Maximilian Witte. The political ideology of conversational AI:  
 608 Converging evidence on ChatGPT’s pro-environmental, left-libertarian orientation, January 2023. URL  
<http://arxiv.org/abs/2301.01768>. arXiv:2301.01768 [cs].

609 Jiwoo Hong, Noah Lee, and James Thorne. Orpo: Monolithic preference optimization without reference model.  
 610 In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 11170–  
 611 11189, 2024.

612 EunJeong Hwang, Bodhisattwa Prasad Majumder, and Niket Tandon. Aligning Language Models to User  
 613 Opinions, May 2023. URL <http://arxiv.org/abs/2305.14929>. arXiv:2305.14929 [cs].

614 Takeshi Kojima, Shixiang (Shane) Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large Lan-  
 615 guage Models are Zero-Shot Reasoners. *Advances in Neural Information Processing Systems*, 35:22199–  
 616 22213, December 2022. URL [https://papers.nips.cc/paper\\_files/paper/2022/hash/8bb0d291acd4acf06ef112099c16f326-Abstract-Conference.html](https://papers.nips.cc/paper_files/paper/2022/hash/8bb0d291acd4acf06ef112099c16f326-Abstract-Conference.html).

617 Lincan Li, Jiaqi Li, Catherine Chen, Fred Gui, Hongjia Yang, Chenxiao Yu, Zhengguang Wang, Jianing Cai,  
 618 Junlong Aaron Zhou, Bolin Shen, Alex Qian, Weixin Chen, Zhongkai Xue, Lichao Sun, Lifang He, Han-  
 619 jie Chen, Kaize Ding, Zijian Du, Fangzhou Mu, Jiaxin Pei, Jieyu Zhao, Swabha Swayamdipta, Willie  
 Neiswanger, Hua Wei, Xiyang Hu, Shixiang Zhu, Tianlong Chen, Yingzhou Lu, Yang Shi, Lianhui Qin,  
 Tianfan Fu, Zhengzhong Tu, Yuzhe Yang, Jaemin Yoo, Jiaheng Zhang, Ryan Rossi, Liang Zhan, Liang Zhao,  
 Emilio Ferrara, Yan Liu, Furong Huang, Xiangliang Zhang, Lawrence Rothenberg, Shuiwang Ji, Philip S.  
 Yu, Yue Zhao, and Yushun Dong. Political-LLM: Large Language Models in Political Science, December  
 620 2024. URL <http://arxiv.org/abs/2412.06864>. arXiv:2412.06864 [cs].

621 Lin Ning, Luyang Liu, Jiaxing Wu, Neo Wu, Devora Berlowitz, Sushant Prakash, Bradley Green, Shawn  
 622 O’Banion, and Jun Xie. User-LLM: Efficient LLM Contextualization with User Embeddings, September  
 623 2024. URL <http://arxiv.org/abs/2402.13598>. arXiv:2402.13598 [cs].

624 OpenAI. Introducing OpenAI o3 and o4-mini, April 2025. URL <https://openai.com/index/introducing-o3-and-o4-mini/>.

625 Joon Sung Park, Carolyn Q. Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb  
 626 Willer, Percy Liang, and Michael S. Bernstein. Generative Agent Simulations of 1,000 People, November  
 627 2024. URL <http://arxiv.org/abs/2411.10109>. arXiv:2411.10109 [cs].

628 Weihong Qi, Hanjia Lyu, and Jiebo Luo. Representation Bias in Political Sample Simulations with Large  
 629 Language Models, July 2024. URL <http://arxiv.org/abs/2407.11409>. arXiv:2407.11409 [cs].

630 Yao Qu and Jue Wang. Performance and biases of Large Language Models in public opinion simulation. *Hu-*  
*631 manities and Social Sciences Communications*, 11(1):1095, August 2024. ISSN 2662-9992. doi: 10.1057/  
 632 s41599-024-03609-x. URL <https://www.nature.com/articles/s41599-024-03609-x>.  
 633 Publisher: Palgrave.

634 Qwen Team. Qwen, April 2025. URL <https://qwenlm.github.io/blog/qwen3/>.

635 Abhinav Rastogi, Albert Q Jiang, Andy Lo, Gabrielle Berrada, Guillaume Lample, Jason Rute, Joep Barmentlo,  
 636 Karmesh Yadav, Kartik Khandelwal, Khyathi Raghavi Chandu, et al. Magistral, 2025.

637 David Rozado. The political preferences of LLMs. *PLOS ONE*, 19(7):e0306621, July 2024. ISSN 1932-6203.  
 638 doi: 10.1371/journal.pone.0306621. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0306621>. Publisher: Public Library of Science.

648 Shibani Santurkar, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. Whose  
 649 Opinions Do Language Models Reflect?, March 2023. URL <http://arxiv.org/abs/2303.17548>.  
 650 arXiv:2303.17548 [cs].

651 John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimiza-  
 652 tion Algorithms, August 2017. URL <http://arxiv.org/abs/1707.06347>. arXiv:1707.06347 [cs].  
 653

654 Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang,  
 655 Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open  
 656 Language Models, April 2024. URL <http://arxiv.org/abs/2402.03300>. arXiv:2402.03300 [cs].

657 Dominik Stammbach, Philine Widmer, Eunjung Cho, Caglar Gulcehre, and Elliott Ash. Aligning Large  
 658 Language Models with Diverse Political Viewpoints, October 2024. URL <http://arxiv.org/abs/2406.14155>. arXiv:2406.14155 [cs].  
 659

660 Seungjong Sun, Eungu Lee, Dongyan Nan, Xiangying Zhao, Wonbyung Lee, Bernard J. Jansen, and Jang Hyun  
 661 Kim. Random Silicon Sampling: Simulating Human Sub-Population Opinion Using a Large Language  
 662 Model Based on Group-Level Demographic Information, February 2024. URL <http://arxiv.org/abs/2402.18144>. arXiv:2402.18144 [cs].  
 663

664 James Thurman and Urs Gasser. Three Case Studies from Switzerland: Smartvote. Technical report, Berkman  
 665 Klein Center, March 2009.

666 Yu-Min Tseng, Yu-Chao Huang, Teng-Yun Hsiao, Wei-Lin Chen, Chao-Wei Huang, Yu Meng, and Yun-Nung  
 667 Chen. Two Tales of Persona in LLMs: A Survey of Role-Playing and Personalization, June 2024. URL  
 668 <https://arxiv.org/abs/2406.01171v3>.

669 Yilei Wang, Jiabao Zhao, Deniz S. Ones, Liang He, and Xin Xu. Evaluating the ability of large language mod-  
 670 els to emulate personality. *Scientific Reports*, 15(1):519, January 2025. ISSN 2045-2322. doi: 10.1038/s41598-024-84109-5.  
 671 URL <https://www.nature.com/articles/s41598-024-84109-5>.  
 672 Publisher: Nature Publishing Group.  
 673

674 Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi,  
 675 Quoc V. Le, and Denny Zhou. Chain-of-Thought Prompting Elicits Reasoning in Large  
 676 Language Models. *Advances in Neural Information Processing Systems*, 35:24824–24837,  
 677 December 2022. URL <https://proceedings.neurips.cc/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html>.  
 678

679 Fengli Xu, Qianyue Hao, Zefang Zong, Jingwei Wang, Yunke Zhang, Jingyi Wang, Xiaochong Lan, Jiahui  
 680 Gong, Tianjian Ouyang, Fanjin Meng, Chenyang Shao, Yuwei Yan, Qinglong Yang, Yiwen Song, Sijian  
 681 Ren, Xinyuan Hu, Yu Li, Jie Feng, Chen Gao, and Yong Li. Towards Large Reasoning Models: A Survey of  
 682 Reinforced Reasoning with Large Language Models, January 2025. URL <http://arxiv.org/abs/2501.09686>. arXiv:2501.09686 [cs].  
 683

684 Chenxiao Yu, Jinyi Ye, Yuangang Li, Zheng Li, Emilio Ferrara, Xiyang Hu, and Yue Zhao. A Large-Scale  
 685 Simulation on Large Language Models for Decision-Making in Political Science, April 2025. URL <http://arxiv.org/abs/2412.15291>. arXiv:2412.15291 [cs].  
 686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702 **A DATASET DETAILS**  
703704 In the following, we provide more information on the datasets used in this work.  
705706 **A.1 SWISS PARTIES**  
707708 Table 4 lists the Swiss parties included in the *smartvote* dataset, along with their political orientations  
709 and group assignments.  
710711 Table 4: The six major Swiss parties and their political positions. All parties' candidates answered  
712 the same 60 questions (48 train / 12 test).  
713

| 714 <b>Party</b> | 715 <b>Political Orientation</b>  | 716 <b>Group</b>  |
|------------------|-----------------------------------|-------------------|
| 715 SVP          | 716 Right-wing conservative       | 717 <b>Right</b>  |
| 716 SP           | 717 Center-left social democratic | 718 <b>Left</b>   |
| 717 FDP          | 718 Center-right liberal          | 719 <b>Right</b>  |
| 718 The Center   | 719 Center/Christian democratic   | 720 <b>Center</b> |
| 719 Green Party  | 720 Left-wing green               | 721 <b>Left</b>   |
| 720 GLP          | 721 Center-left green liberal     | 722 <b>Center</b> |

722 **A.2 WAHL-O-MAT**  
723724 Table 5 lists the German parties included in the *Wahl-o-Mat* dataset, along with their political orientations, group assignments, as well as the train and test set statistics.  
725726 Table 5: Number of questions per party in the German dataset across all elections (2021–2025). The  
727 test sets contain 30 questions for all parties. A → Yes, B → No, C → Neutral.  
728

| 729 <b>Party</b> | 730 <b>Political Orientation</b>      | 731 <b># Train Q.</b> | 732 <b>Train. distr.</b>  | 733 <b>Test distr.</b>   | 734 <b>Group</b>  |
|------------------|---------------------------------------|-----------------------|---------------------------|--------------------------|-------------------|
|                  |                                       |                       | 735 Yes/Neutral/No        | 736 Yes/Neutral/No       |                   |
| 732 CDU/CSU      | 733 Center-right Christian democratic | 734 646               | 735 52.32 / 39.16 / 8.51  | 736 60.0 / 36.67 / 3.33  | 737 <b>Right</b>  |
| 733 SPD          | 734 Center-left social democratic     | 735 760               | 736 55.92 / 32.63 / 11.45 | 737 63.33 / 33.33 / 3.33 | 738 <b>Center</b> |
| 734 Grüne        | 735 Center-left green                 | 736 722               | 737 52.08 / 37.40 / 10.53 | 738 73.33 / 20.0 / 6.67  | 739 <b>Left</b>   |
| 735 FDP          | 736 Center-right liberal              | 737 760               | 738 48.68 / 42.37 / 8.95  | 739 46.67 / 53.33 / 0.00 | 740 <b>Center</b> |
| 736 Die Linke    | 737 Left-wing socialist               | 738 646               | 739 52.17 / 41.80 / 6.04  | 740 56.67 / 40.0 / 3.33  | 741 <b>Left</b>   |
| 737 AfD          | 738 Far-right populist                | 739 722               | 740 45.43 / 48.20 / 6.37  | 741 30.0 / 66.67 / 3.33  | 742 <b>Right</b>  |

738 **A.3 ANES**  
739740 Table 6 provides the assignment of ANES ideologies to groups.  
741742 Table 6: Assignment of ANES ideology self-placement answers to groups.  
743

| 744 <b>Ideology</b>        | 745 <b>Group</b>  |
|----------------------------|-------------------|
| 745 Extremely liberal      | 746 <b>Left</b>   |
| 746 Liberal                | 747 <b>Left</b>   |
| 747 Slightly liberal       | 748 <b>Center</b> |
| 748 Moderate               | 749 <b>Center</b> |
| 749 Slightly conservative  | 750 <b>Center</b> |
| 750 Conservative           | 751 <b>Right</b>  |
| 751 Extremely conservative | 752 <b>Right</b>  |

753 **A.4 PROMPTS FOR SYNTHETIC DATASETS**  
754

755 The following prompt was used to generate the synthetic argument datasets.

756 You are a [politician / voter] in [COUNTRY]. You reply  
 757 in English only. We present you with a political  
 758 issue or question, and you provide us with three to  
 759 five diverse yet concise arguments [for / against]  
 760 this issue or question. [ISSUE]

761 For *smartvote* and *ANES*, COUNTRY was replaced with Switzerland and the USA, respectively.  
 762 Similarly, politician was used for *smartvote*, while voter was used for *ANES*. For the biased  
 763 datasets used in Section 4.8, the *smartvote* prompt was changed to the following:  
 764

765 You are a politician in Switzerland. You reply in  
 766 English only. We present you with a political issue  
 767 or question, and you provide us with a [progressive /  
 768 conservative] argument [for / against] this issue or  
 769 question. [ISSUE]

## 771 B ADDITIONAL REASONING TRACES

773 In most cases, the reasoning traces are on-topic, align with the final answer, and are internally  
 774 consistent. In Table 7, we present an example from a German party in the *Left* group (SPD). Here,  
 775 the agent is also making use of information it acquired, most likely during the SFT stage on the  
 776 training set.

778 Table 7: Example of successful reasoning. The reasoning trace and the final answer are consistent.  
 779 The agent is also making use of additional information about the party position, most likely acquired  
 780 during the SFT stage on the training set.

---

782 **Question:** Social programs should be maintained even at the cost of higher taxes

783 <reasoning> We are committed to social justice and therefore to strong social work and high social  
 784 spending. For us, social policy always means more than mere benefit payments. We are committed to free  
 785 childcare for all children up to the age of ten years and for senior citizens and nursing services that look after  
 786 them with respect and dignity. </reasoning>

787 <answer>A</answer>

---

788  
 789 Agents sometimes resort to rather unconventional arguments. For instance, in the trace in Table 8,  
 790 an agent trained on a left-wing politician from Switzerland suggested that non-proliferation could  
 791 suffer if Switzerland were to build new nuclear power plants. While the argument and the reasoning  
 792 are consistent, it is rather unlikely that this argument would be made.

794 Table 8: Example of a reasoning trace where the reasoning is formally correct, but the argument  
 795 made is unlikely to be made in this context.

---

798 **Question:** Should the construction of new nuclear power plants in Switzerland be allowed again?

799 <reasoning> Switzerland has historically been committed to a policy of non-proliferation of nuclear  
 800 weapons. Allowing new nuclear power plants could potentially compromise this policy, as there is a risk that  
 801 nuclear materials or technology could be diverted or transferred to other countries with questionable  
 802 intentions, potentially contributing to the development of nuclear weapons.</reasoning>

803 <answer>B</answer>

---

804 Similarly, in Table 9, the agent starts in the right direction for the unit it is trained to represent, but  
 805 then switches topics to prostitution.

## 807 C ADDITIONAL FIGURES

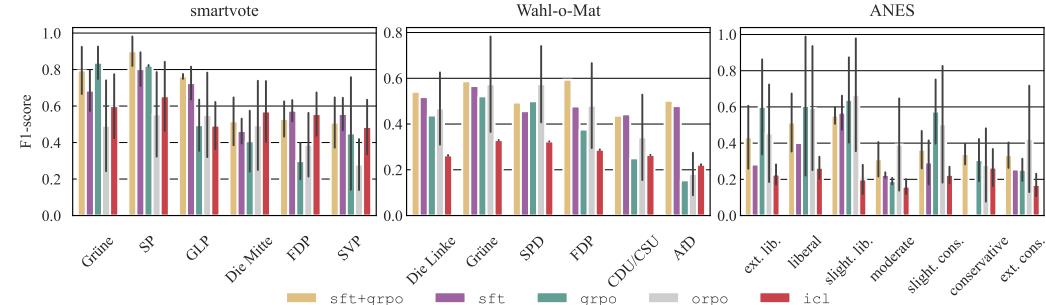
809 Figure 8 provides per-party F1 scores, similar to Fig. 3.

810 Table 9: Example of a trace where the model goes off-topic. This is most likely prompted by the  
 811 presence of the word “sex”.

813 **Question:** The legalization of same-sex marriages is a good thing.

814 <reasoning> In the future, there should be no further debates on the equality of homo and heterosexual  
 815 couples. There must also be further prostitution-free areas to protect the neighborhood from the influence of  
 816 street prostitution. </reasoning>

817 <answer>A</answer>



832 Figure 8: Same data as Fig. 3, broken down by party. The party-level data shows a similar trend,  
 833 SFT+GRPO generally outperforms SFT and GRPO. F1 tends to be better on left and center-left  
 834 parties, except the German center-right FDP.

## 836 D ANES RECODING SCHEMES

839 Questions that ask for (dis-)agreement on a Likert scale were left as is, and the answer options are  
 840 mapped such that all agreeing options become “Yes” or “Agree”, and all disagreeing options become  
 841 “No” or “Disagree”. Options like “Neither agree nor disagree” or “None of the above” are mapped  
 842 to neutral. All other types of questions were rewritten to allow for this answer style. For instance,  
 843 consider the following original ANES question:

844 When women demand equality these days, How often are  
 845 they actually seeking special favors?  
 846 1. Always  
 847 2. Most of the time  
 848 3. About half the time  
 849 4. Some of the time  
 850 5. Never

851 This question can be rewritten as a statement like:

853 When women demand equality these days, they are  
 854 actually seeking special favors.

856 For recoding the answers, we consider two schemes:

- 858 • **Conservative shceme:** The original options 1 and 2 become Yes, option 3 becomes  
 859 Neutral, and options 4 and 5 are mapped to No. Generally, options in the middle  
 860 of the scale, like moderately important, a moderate amount, about half  
 861 of the time, etc., are mapped to Neutral under this scheme.
- 862 • **Aggressive scheme:** Options 1 to 4 are mapped to Yes and only option 5 is mapped to No.  
 863 Generally, only clearly negative options are mapped to No under this scheme. There is no  
 Neutral option, unless the original options contain one.

864 The conservative scheme treats only clearly positive frequency responses as agreement, while the  
 865 aggressive scheme treats any non-zero frequency as agreement.  
 866

867 **D.1 FURTHER EXAMPLES**  
 868

869 The question  
 870

871 How often do Members of Congress change their votes  
 872 on legislation because someone donates money to their  
 873 campaign?  
 874 1. Never  
 875 2. Rarely  
 876 3. A moderate amount of time  
 877 4. Very often  
 878 5. All the time

879 becomes  
 880

881 Do Members of Congress change their votes on  
 882 legislation because someone donates money to their  
 883 campaign?''

884 with mappings (4, 5) → Yes , (3) → Neutral, (1, 2) → No using the conservative scheme, and  
 885 (2,3,4,5) → Yes and (1) → No using the aggressive scheme. As an other example, the question  
 886

887 Does the increasing number of people of many different  
 888 races and ethnic groups in the United States make this  
 889 country a better place to live, a worse place to live,  
 890 or does it make no difference?  
 891 1. Better  
 892 2. Worse  
 893 3. Makes no difference

894 becomes  
 895

896 Does the increasing number of people of many different  
 897 races and ethnic groups in the United States make this  
 898 country a better place?  
 899

900 with mappings (1) → Yes, (2) → No and (3) → Neutral in both recoding schemes.  
 901

902 **E ADDITIONAL MATERIAL**  
 903

904 The following sections provide additional material to contextualize the findings in the main text.  
 905

906 **E.1 ADDITIONAL RESULT TABLES**  
 907

908 Table 10 presents accuracy scores across models and configurations. Similarly, in Table 11, we  
 909 present accuracy scores without the Neutral class. In Table 12 and Table 13 , we report one-  
 910 tailed Welch's t-tests (SFT+GRPO > baseline) and Cohen's  $d$  effect sizes for the results in Table 3  
 911 (macro-F1) and Table 10 (mean accuracy), respectively.  
 912

913 **E.2 IMPACT OF RECODING ANES**  
 914

915 Since the recoding is not unique and can potentially lead to information loss and semantic shifts,  
 916 we investigate the impact of different recoding schemes on the political groups present in the *ANES*  
 917 dataset. Details on the different schemes are provided in Appendix D. The key difference lies in  
 how moderate responses are treated: the conservative scheme maps them to neutral/disagreement,

918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
9999

Table 10: **Mean accuracies (%)** with standard deviations, averaged over 8 stochastic decoding runs at temperature  $T=1.0$ . For each run, we first compute the per-unit accuracy (candidate/party/respondent) over that unit’s test items and then average across units. The table reports the across-run mean  $\pm$  s.d. Baselines: `random` (uniform; 50.0/33.3 on binary/ternary labels by construction) and `majority` (deterministic). **Bold** indicates the best method per base model and dataset, with significance reported in Table 13.

| 925<br>926<br>927<br>928<br>929<br>930<br>931<br>932<br>933<br>934<br>935<br>936<br>937<br>938<br>939<br>940<br>941<br>942<br>943<br>944<br>945<br>946<br>947<br>948<br>949<br>950<br>951<br>952<br>953<br>954<br>955<br>956<br>957<br>958<br>959<br>960<br>961<br>962<br>963<br>964<br>965<br>966<br>967<br>968<br>969<br>970<br>971<br>972<br>973<br>974<br>975<br>976<br>977<br>978<br>979<br>9799 | 925<br>926<br>927<br>928<br>929<br>930<br>931<br>932<br>933<br>934<br>935<br>936<br>937<br>938<br>939<br>940<br>941<br>942<br>943<br>944<br>945<br>946<br>947<br>948<br>949<br>950<br>951<br>952<br>953<br>954<br>955<br>956<br>957<br>958<br>959<br>960<br>961<br>962<br>963<br>964<br>965<br>966<br>967<br>968<br>969<br>970<br>971<br>972<br>973<br>974<br>975<br>976<br>977<br>978<br>979<br>9799 | 925<br>926<br>927<br>928<br>929<br>930<br>931<br>932<br>933<br>934<br>935<br>936<br>937<br>938<br>939<br>940<br>941<br>942<br>943<br>944<br>945<br>946<br>947<br>948<br>949<br>950<br>951<br>952<br>953<br>954<br>955<br>956<br>957<br>958<br>959<br>960<br>961<br>962<br>963<br>964<br>965<br>966<br>967<br>968<br>969<br>970<br>971<br>972<br>973<br>974<br>975<br>976<br>977<br>978<br>979<br>9799 | Dataset   |                                    |                                    |                                    |  |  |  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|------------------------------------|------------------------------------|------------------------------------|--|--|--|
|                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       | smartvote | WoM                                | ANES                               |                                    |  |  |  |
| 925<br>926<br>927<br>928<br>929<br>930<br>931<br>932<br>933<br>934<br>935<br>936<br>937<br>938<br>939<br>940<br>941<br>942<br>943<br>944<br>945<br>946<br>947<br>948<br>949<br>950<br>951<br>952<br>953<br>954<br>955<br>956<br>957<br>958<br>959<br>960<br>961<br>962<br>963<br>964<br>965<br>966<br>967<br>968<br>969<br>970<br>971<br>972<br>973<br>974<br>975<br>976<br>977<br>978<br>979<br>9799 | 925<br>926<br>927<br>928<br>929<br>930<br>931<br>932<br>933<br>934<br>935<br>936<br>937<br>938<br>939<br>940<br>941<br>942<br>943<br>944<br>945<br>946<br>947<br>948<br>949<br>950<br>951<br>952<br>953<br>954<br>955<br>956<br>957<br>958<br>959<br>960<br>961<br>962<br>963<br>964<br>965<br>966<br>967<br>968<br>969<br>970<br>971<br>972<br>973<br>974<br>975<br>976<br>977<br>978<br>979<br>9799 | 925<br>926<br>927<br>928<br>929<br>930<br>931<br>932<br>933<br>934<br>935<br>936<br>937<br>938<br>939<br>940<br>941<br>942<br>943<br>944<br>945<br>946<br>947<br>948<br>949<br>950<br>951<br>952<br>953<br>954<br>955<br>956<br>957<br>958<br>959<br>960<br>961<br>962<br>963<br>964<br>965<br>966<br>967<br>968<br>969<br>970<br>971<br>972<br>973<br>974<br>975<br>976<br>977<br>978<br>979<br>9799 | random    | 50.00                              | 33.33                              | 33.33                              |  |  |  |
|                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       | majority  | 37.43                              | 27.44                              | 22.98                              |  |  |  |
| 925<br>926<br>927<br>928<br>929<br>930<br>931<br>932<br>933<br>934<br>935<br>936<br>937<br>938<br>939<br>940<br>941<br>942<br>943<br>944<br>945<br>946<br>947<br>948<br>949<br>950<br>951<br>952<br>953<br>954<br>955<br>956<br>957<br>958<br>959<br>960<br>961<br>962<br>963<br>964<br>965<br>966<br>967<br>968<br>969<br>970<br>971<br>972<br>973<br>974<br>975<br>976<br>977<br>978<br>979<br>9799 | 925<br>926<br>927<br>928<br>929<br>930<br>931<br>932<br>933<br>934<br>935<br>936<br>937<br>938<br>939<br>940<br>941<br>942<br>943<br>944<br>945<br>946<br>947<br>948<br>949<br>950<br>951<br>952<br>953<br>954<br>955<br>956<br>957<br>958<br>959<br>960<br>961<br>962<br>963<br>964<br>965<br>966<br>967<br>968<br>969<br>970<br>971<br>972<br>973<br>974<br>975<br>976<br>977<br>978<br>979<br>9799 | 925<br>926<br>927<br>928<br>929<br>930<br>931<br>932<br>933<br>934<br>935<br>936<br>937<br>938<br>939<br>940<br>941<br>942<br>943<br>944<br>945<br>946<br>947<br>948<br>949<br>950<br>951<br>952<br>953<br>954<br>955<br>956<br>957<br>958<br>959<br>960<br>961<br>962<br>963<br>964<br>965<br>966<br>967<br>968<br>969<br>970<br>971<br>972<br>973<br>974<br>975<br>976<br>977<br>978<br>979<br>9799 | icl       | 63.91 $\pm$ 3.79                   | 31.44 $\pm$ 0.02                   | 44.39 $\pm$ 1.43                   |  |  |  |
|                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       | ORPO      | 59.23 $\pm$ 2.47                   | 57.48 $\pm$ 5.97                   | 44.15 $\pm$ 5.61                   |  |  |  |
|                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       | SFT       | 67.23 $\pm$ 1.00                   | 68.25 $\pm$ 4.30                   | 58.82 $\pm$ 0.93                   |  |  |  |
|                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       | GRPO      | 66.44 $\pm$ 0.86                   | 60.63 $\pm$ 1.50                   | 52.35 $\pm$ 1.35                   |  |  |  |
|                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       | SFT+GRPO  | <b>70.53 <math>\pm</math> 1.51</b> | <b>75.06 <math>\pm</math> 3.30</b> | <b>59.21 <math>\pm</math> 0.62</b> |  |  |  |
| 925<br>926<br>927<br>928<br>929<br>930<br>931<br>932<br>933<br>934<br>935<br>936<br>937<br>938<br>939<br>940<br>941<br>942<br>943<br>944<br>945<br>946<br>947<br>948<br>949<br>950<br>951<br>952<br>953<br>954<br>955<br>956<br>957<br>958<br>959<br>960<br>961<br>962<br>963<br>964<br>965<br>966<br>967<br>968<br>969<br>970<br>971<br>972<br>973<br>974<br>975<br>976<br>977<br>978<br>979<br>9799 | 925<br>926<br>927<br>928<br>929<br>930<br>931<br>932<br>933<br>934<br>935<br>936<br>937<br>938<br>939<br>940<br>941<br>942<br>943<br>944<br>945<br>946<br>947<br>948<br>949<br>950<br>951<br>952<br>953<br>954<br>955<br>956<br>957<br>958<br>959<br>960<br>961<br>962<br>963<br>964<br>965<br>966<br>967<br>968<br>969<br>970<br>971<br>972<br>973<br>974<br>975<br>976<br>977<br>978<br>979<br>9799 | 925<br>926<br>927<br>928<br>929<br>930<br>931<br>932<br>933<br>934<br>935<br>936<br>937<br>938<br>939<br>940<br>941<br>942<br>943<br>944<br>945<br>946<br>947<br>948<br>949<br>950<br>951<br>952<br>953<br>954<br>955<br>956<br>957<br>958<br>959<br>960<br>961<br>962<br>963<br>964<br>965<br>966<br>967<br>968<br>969<br>970<br>971<br>972<br>973<br>974<br>975<br>976<br>977<br>978<br>979<br>9799 | icl       | 66.09 $\pm$ 2.85                   | 48.67 $\pm$ 0.73                   | 41.32 $\pm$ 2.98                   |  |  |  |
|                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       | ORPO      | 39.02 $\pm$ 2.04                   | 38.56 $\pm$ 3.51                   | 34.07 $\pm$ 2.08                   |  |  |  |
|                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       | SFT       | 65.18 $\pm$ 1.91                   | 61.74 $\pm$ 4.63                   | 47.01 $\pm$ 1.79                   |  |  |  |
|                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       | GRPO      | 67.04 $\pm$ 1.69                   | 53.19 $\pm$ 2.33                   | 52.92 $\pm$ 0.72                   |  |  |  |
|                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       | SFT+GRPO  | <b>71.27 <math>\pm</math> 2.83</b> | <b>71.16 <math>\pm</math> 1.60</b> | <b>55.14 <math>\pm</math> 0.41</b> |  |  |  |
| 925<br>926<br>927<br>928<br>929<br>930<br>931<br>932<br>933<br>934<br>935<br>936<br>937<br>938<br>939<br>940<br>941<br>942<br>943<br>944<br>945<br>946<br>947<br>948<br>949<br>950<br>951<br>952<br>953<br>954<br>955<br>956<br>957<br>958<br>959<br>960<br>961<br>962<br>963<br>964<br>965<br>966<br>967<br>968<br>969<br>970<br>971<br>972<br>973<br>974<br>975<br>976<br>977<br>978<br>979<br>9799 | 925<br>926<br>927<br>928<br>929<br>930<br>931<br>932<br>933<br>934<br>935<br>936<br>937<br>938<br>939<br>940<br>941<br>942<br>943<br>944<br>945<br>946<br>947<br>948<br>949<br>950<br>951<br>952<br>953<br>954<br>955<br>956<br>957<br>958<br>959<br>960<br>961<br>962<br>963<br>964<br>965<br>966<br>967<br>968<br>969<br>970<br>971<br>972<br>973<br>974<br>975<br>976<br>977<br>978<br>979<br>9799 | 925<br>926<br>927<br>928<br>929<br>930<br>931<br>932<br>933<br>934<br>935<br>936<br>937<br>938<br>939<br>940<br>941<br>942<br>943<br>944<br>945<br>946<br>947<br>948<br>949<br>950<br>951<br>952<br>953<br>954<br>955<br>956<br>957<br>958<br>959<br>960<br>961<br>962<br>963<br>964<br>965<br>966<br>967<br>968<br>969<br>970<br>971<br>972<br>973<br>974<br>975<br>976<br>977<br>978<br>979<br>9799 | icl       | 71.41 $\pm$ 3.54                   | 44.64 $\pm$ 0.67                   | 42.46 $\pm$ 5.13                   |  |  |  |
|                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       | ORPO      | 35.93 $\pm$ 3.66                   | 36.10 $\pm$ 5.18                   | 31.32 $\pm$ 2.23                   |  |  |  |
|                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       | SFT       | 70.83 $\pm$ 1.58                   | 72.56 $\pm$ 2.40                   | 52.80 $\pm$ 0.94                   |  |  |  |
|                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       | GRPO      | 68.44 $\pm$ 1.23                   | 72.05 $\pm$ 2.99                   | 62.33 $\pm$ 0.51                   |  |  |  |
|                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                       | SFT+GRPO  | <b>73.92 <math>\pm</math> 1.89</b> | <b>75.10 <math>\pm</math> 2.80</b> | <b>62.33 <math>\pm</math> 0.78</b> |  |  |  |

while the aggressive scheme treats most non-zero responses as agreement. We train Llama 3.1 8B on the two recoding schemes and find that mean F1-scores on the aggressive scheme are higher than on the conservative one (45.33  $\pm$  1.37 vs. 40.67  $\pm$  0.93 reported in Table 3). The confusion matrices in Fig. 9 indicate that the improvement is mostly driven by a lower number of *Neutrals* in the aggressive scheme. These results highlight that recoding choices can substantially impact both model performance and the political diversity represented in simulated populations.

Figure 9: Confusion matrices for Llama 3.1 8B (SFT+GRPO) trained on ANES with a more aggressive recoding scheme. A  $\rightarrow$  Yes, B  $\rightarrow$  No, C  $\rightarrow$  Neutral.

### E.3 REGRESSION TABLES

Table 14 provides the regression table for the regression shown in 4. Similarly, Table 15 provides the regression table for the corresponding analysis shown in Fig. 10.

### E.4 IMPACT OF THE NUMBER OF TRAINING QUESTIONS

We ablate the number of questions in the WoM training set for the SFT+GRPO configuration. The results for Llama 3.1 8B are presented in Fig. 11. Increasing the number of training questions is beneficial for all parties, except the center-right FDP party. The FDP’s performance may be explained by a shift in the party’s positions due to dwindling voter numbers between 2021 and 2025. The largest increase in accuracy is observed for the far-right AfD, which we attribute to the fact that

972  
 973 Table 11: **Mean accuracies** (%) with standard deviations when **Neutral** is removed. Results are  
 974 the averages over 8 stochastic decoding runs at temperature  $T=1.0$ . For each run, we first com-  
 975 pute the per-unit accuracy (candidate/party/respondent) over that unit’s test items and then average  
 976 across units. The table reports the across-run mean  $\pm$  s.d. Baselines: **random** (uniform; 50.0/33.3  
 977 on binary/ternary labels by construction) and **majority** (deterministic). **Bold** indicates the best  
 978 method per base model and dataset (higher is better). *smartvote* results are unchanged from Table 10  
 979 because there are no neutral answers on *smartvote*.

| 980 <b>Base model</b>   | 981 <b>Method</b>    | 982 <b>Dataset</b>    |                       |                       |
|-------------------------|----------------------|-----------------------|-----------------------|-----------------------|
|                         |                      | 983 <i>smartvote</i>  | 984 <i>WoM</i>        | 985 <i>ANES</i>       |
| 986 Untrained baselines | 987 <b>random</b>    | 988 50.00             | 989 33.33             | 990 33.33             |
|                         | 991 <b>majority</b>  | 992 37.43             | 993 27.44             | 994 22.98             |
| 995 Llama 3.1 8B        | 996 <b>ic1</b>       | 997 $63.91 \pm 3.79$  | 998 $31.03 \pm 0.19$  | 999 $47.15 \pm 1.69$  |
|                         | 999 <b>ORPO</b>      | 1000 $59.23 \pm 2.47$ | 1001 $59.28 \pm 6.15$ | 1002 $44.15 \pm 5.61$ |
|                         | 1002 <b>SFT</b>      | 1003 $67.23 \pm 0.10$ | 1004 $70.11 \pm 4.36$ | 1005 $62.47 \pm 0.92$ |
|                         | 1005 <b>GRPO</b>     | 1006 $66.44 \pm 0.86$ | 1007 $62.93 \pm 1.54$ | 1008 $64.15 \pm 1.51$ |
|                         | 1008 <b>SFT+GRPO</b> | 1009 $70.53 \pm 1.51$ | 1010 $78.19 \pm 3.38$ | 1011 $71.35 \pm 0.80$ |
| 1012 Qwen3 8B           | 1013 <b>ic1</b>      | 1014 $66.09 \pm 2.85$ | 1015 $31.03 \pm 0.19$ | 1016 $19.11 \pm 2.93$ |
|                         | 1016 <b>ORPO</b>     | 1017 $39.02 \pm 2.04$ | 1018 $37.98 \pm 3.29$ | 1019 $37.04 \pm 1.49$ |
|                         | 1019 <b>SFT</b>      | 1020 $65.18 \pm 1.91$ | 1021 $63.62 \pm 4.74$ | 1022 $54.38 \pm 2.03$ |
|                         | 1022 <b>GRPO</b>     | 1023 $67.04 \pm 1.69$ | 1024 $55.30 \pm 2.44$ | 1025 $65.38 \pm 0.91$ |
|                         | 1025 <b>SFT+GRPO</b> | 1026 $71.27 \pm 2.83$ | 1027 $73.69 \pm 1.60$ | 1028 $67.55 \pm 0.72$ |
| 1029 Magistral 24B      | 1030 <b>ic1</b>      | 1031 $66.09 \pm 2.85$ | 1032 $46.62 \pm 0.80$ | 1033 $46.27 \pm 5.94$ |
|                         | 1033 <b>ORPO</b>     | 1034 $35.93 \pm 3.66$ | 1035 $36.39 \pm 5.38$ | 1036 $33.67 \pm 1.86$ |
|                         | 1036 <b>SFT</b>      | 1037 $70.83 \pm 1.58$ | 1038 $74.71 \pm 2.71$ | 1039 $62.67 \pm 1.17$ |
|                         | 1039 <b>GRPO</b>     | 1040 $68.44 \pm 1.23$ | 1041 $77.21 \pm 2.72$ | 1042 $66.34 \pm 0.55$ |
|                         | 1042 <b>SFT+GRPO</b> | 1043 $73.92 \pm 1.89$ | 1044 $74.71 \pm 2.71$ | 1045 $69.99 \pm 0.85$ |

999 Table 12: **Statistical significance and effect sizes for SFT+GRPO improvements in macro-F1.**  
 1000 Results correspond to the experiments reported in Table 3. We perform one-tailed Welch’s t-tests  
 1001 (SFT+GRPO > baseline) and report Cohen’s  $d$  effect sizes. **Green**: significant after Bonferroni  
 1002 correction ( $p < 0.0014$  for 36 tests). **Blue**: significant without correction ( $p < 0.05$ ). **Red**: not  
 1003 significant.

| 1004 <b>Model</b>  | 1005 <b>Baseline</b> | 1006 <b>P-values</b>  |                    |                    | 1007 <b>Cohen’s d</b> |                   |                   |
|--------------------|----------------------|-----------------------|--------------------|--------------------|-----------------------|-------------------|-------------------|
|                    |                      | 1008 <i>smartvote</i> | 1009 <i>WoM</i>    | 1010 <i>ANES</i>   | 1011 <i>smartvote</i> | 1012 <i>WoM</i>   | 1013 <i>ANES</i>  |
| 1014 Llama 3.1 8B  | 1015 <b>ic1</b>      | 1016 <b>0.0003</b>    | 1017 <b>0.0000</b> | 1018 <b>0.0000</b> | 1019 <b>2.54</b>      | 1020 <b>8.51</b>  | 1021 <b>12.30</b> |
|                    | 1021 <b>ORPO</b>     | 1022 <b>0.0000</b>    | 1023 <b>0.0008</b> | 1024 <b>0.0093</b> | 1025 <b>9.29</b>      | 1026 <b>1.99</b>  | 1027 <b>1.50</b>  |
|                    | 1027 <b>SFT</b>      | 1028 <b>0.0013</b>    | 1029 <b>0.0409</b> | 1030 <b>0.9991</b> | 1031 <b>1.90</b>      | 1032 <b>0.94</b>  | 1033 <b>-1.95</b> |
|                    | 1033 <b>GRPO</b>     | 1034 <b>0.0000</b>    | 1035 <b>0.0000</b> | 1036 <b>0.0000</b> | 1037 <b>6.61</b>      | 1038 <b>4.74</b>  | 1039 <b>5.39</b>  |
| 1040 Qwen3 8B      | 1041 <b>ic1</b>      | 1042 <b>0.0140</b>    | 1043 <b>0.0000</b> | 1044 <b>0.0000</b> | 1045 <b>1.23</b>      | 1046 <b>16.72</b> | 1047 <b>11.75</b> |
|                    | 1047 <b>ORPO</b>     | 1048 <b>0.0000</b>    | 1049 <b>0.0000</b> | 1050 <b>0.0000</b> | 1051 <b>14.87</b>     | 1052 <b>8.39</b>  | 1053 <b>11.46</b> |
|                    | 1053 <b>SFT</b>      | 1054 <b>0.0099</b>    | 1055 <b>0.0004</b> | 1056 <b>0.0004</b> | 1057 <b>1.32</b>      | 1058 <b>2.32</b>  | 1059 <b>2.67</b>  |
|                    | 1059 <b>GRPO</b>     | 1060 <b>0.0037</b>    | 1061 <b>0.0000</b> | 1062 <b>0.0000</b> | 1063 <b>1.63</b>      | 1064 <b>10.50</b> | 1065 <b>8.70</b>  |
| 1066 Magistral 24B | 1067 <b>ic1</b>      | 1068 <b>0.0102</b>    | 1069 <b>0.0000</b> | 1070 <b>0.0000</b> | 1071 <b>1.36</b>      | 1072 <b>11.91</b> | 1073 <b>11.90</b> |
|                    | 1073 <b>ORPO</b>     | 1074 <b>0.0000</b>    | 1075 <b>0.0000</b> | 1076 <b>0.0000</b> | 1077 <b>21.95</b>     | 1078 <b>8.79</b>  | 1079 <b>11.81</b> |
|                    | 1079 <b>SFT</b>      | 1080 <b>0.0049</b>    | 1081 <b>0.1842</b> | 1082 <b>0.0000</b> | 1083 <b>1.50</b>      | 1084 <b>0.47</b>  | 1085 <b>6.88</b>  |
|                    | 1085 <b>GRPO</b>     | 1086 <b>0.0000</b>    | 1087 <b>0.0909</b> | 1088 <b>0.0046</b> | 1089 <b>4.90</b>      | 1090 <b>0.70</b>  | 1091 <b>1.51</b>  |

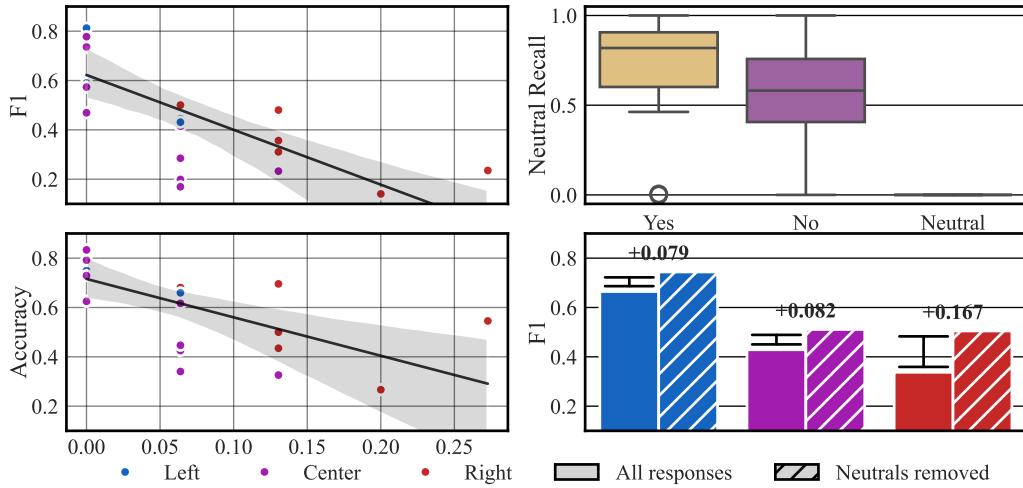
1092 their positions are furthest from the LLM’s position, and SFT thus has an aligning or “corrective”  
 1093 effect which becomes stronger as more data is added.

## 1094 E.5 PCA DETAILS

1095 To obtain the results shown in Fig. 2 and Fig. 12, we compute the PCA as follows: We use the  
 1096 complete *smartvote* dataset for the 2023 national elections across all major parties and all candidates.

1026 **Table 13: Statistical significance and effect sizes for SFT+GRPO improvements in accuracy.**  
1027 Results correspond to the experiments reported in Table 10. We perform one-tailed Welch’s t-tests  
1028 (SFT+GRPO  $>$  baseline) and report Cohen’s  $d$  effect sizes. **Green:** significant after Bonferroni  
1029 correction ( $p < 0.0014$  for 36 tests). **Blue:** significant without correction ( $p < 0.05$ ). **Red:** not  
1030 significant.

| 1032 | Model         | Baseline | P-values  |        |        | Cohen’s d |       |       |
|------|---------------|----------|-----------|--------|--------|-----------|-------|-------|
|      |               |          | smartvote | WoM    | ANES   | smartvote | WoM   | ANES  |
| 1034 | Llama 3.1 8B  | ic1      | 0.0006    | 0.0000 | 0.0000 | 2.29      | 18.66 | 13.45 |
|      |               | ORPO     | 0.0000    | 0.0000 | 0.0001 | 5.52      | 3.64  | 3.77  |
|      |               | SFT      | 0.0001    | 0.0017 | 0.1714 | 2.58      | 1.78  | 0.49  |
|      |               | GRPO     | 0.0000    | 0.0000 | 0.0000 | 3.33      | 5.63  | 6.53  |
| 1038 | Qwen3 8B      | ic1      | 0.0013    | 0.0000 | 0.0000 | 1.82      | 18.09 | 6.50  |
|      |               | ORPO     | 0.0000    | 0.0000 | 0.0000 | 13.07     | 11.95 | 14.06 |
|      |               | SFT      | 0.0001    | 0.0002 | 0.0000 | 2.52      | 2.72  | 6.26  |
|      |               | GRPO     | 0.0019    | 0.0000 | 0.0000 | 1.81      | 8.99  | 3.79  |
| 1042 | Magistral 24B | ic1      | 0.0527    | 0.0000 | 0.0000 | 0.88      | 14.96 | 5.42  |
|      |               | ORPO     | 0.0000    | 0.0000 | 0.0000 | 13.04     | 9.37  | 18.56 |
|      |               | SFT      | 0.0017    | 0.0361 | 0.0000 | 1.77      | 0.97  | 11.03 |
|      |               | GRPO     | 0.0000    | 0.0269 | 0.5000 | 3.44      | 1.05  | 0.00  |



1064 **Figure 10: F1 score (A) and accuracy (B) vs. Neutral response rate.** Predictive performance  
1065 decreases significantly as Neutral response rate increases. Comparing the results presented here  
1066 to those in Fig. 4, we observe the same trend regardless of the coding scheme. (D): Performance in  
1067 terms of F1 score on ANES before (left) and after (right) removing Neutral instances. All groups  
1068 improve, but **Center** only marginally. The gap between **Center** and **Right** closes. Performance in  
1069 predicting Neutral decreased (C). Results were obtained for Llama 3.1 8B using the SFT+GRPO  
1070 configuration.

1071  
1072  
1073 We convert the original 4-point Likert scale answers to yes-no, as we did for the *smartvote* subset  
1074 studied in the remainder of the paper. Yes answers are subsequently mapped to 1, and No answers  
1075 to 0. We then compute the PCA of the resulting matrix, where the rows correspond to candidates,  
1076 and the columns correspond to questions.

1077 For agents, we replace the ground-truth test set answers with the agents’ predictions, and project the  
1078 resulting vectors onto the first two principal components. Similarly, for the inverted positions shown  
1079 in Fig. 12, we invert all ground-truth answers and project the resulting vectors.

1080 Table 14: Regression table for the analysis presented in Fig. 4 and Section 4.6: Model performance  
 1081 vs. Neutral base rate.

1082

| Metric           | N       | Intercept ( $\alpha$ ) | Slope ( $\beta$ ) | Slope SE | Slope 95% CI       |
|------------------|---------|------------------------|-------------------|----------|--------------------|
| F1 Score (Macro) | 21      | 0.5517                 | -0.6741           | 0.2091   | [-1.3039, -0.4084] |
| Accuracy         | 21      | 0.7803                 | -0.8562           | 0.2139   | [-1.1117, -0.2364] |
| Metric           | r       | $R^2$                  | p-value           | RMSE     |                    |
| F1 Score (Macro) | -0.5946 | 0.3535                 | 0.004472          | 0.1250   |                    |
| Accuracy         | -0.6763 | 0.4574                 | < 0.001           | 0.1279   |                    |

1090

1091 Table 15: Regression table for the analysis presented in Fig. 10: Model performance vs. Neutral  
 1092 base rate.

1093

| Metric           | N       | Intercept ( $\alpha$ ) | Slope ( $\beta$ ) | Slope SE | Slope 95% CI       |
|------------------|---------|------------------------|-------------------|----------|--------------------|
| F1 Score (Macro) | 21      | 0.6225                 | -2.2231           | 0.5049   | [-3.2799, -1.1662] |
| Accuracy         | 21      | 0.7157                 | -1.5571           | 0.4496   | [-2.4981, -0.6161] |
| Metric           | r       | $R^2$                  | p-value           | RMSE     |                    |
| F1 Score (Macro) | -0.7106 | 0.5050                 | < 0.001           | 0.1608   |                    |
| Accuracy         | -0.6221 | 0.3870                 | 0.002603          | 0.1431   |                    |

1101

1102

## E.6 INVERSION

1104

1105

Figure 12 shows candidate positions before and after flipping all answer options. As can be seen, this operation roughly corresponds to a rotation by  $180^\circ$  about the origin.

1106

1107

## F TRAINING DETAILS

1109

1110

In the following, we provide information on training settings and hyperparameters.

1111

1112

**System prompt** . We used the following system prompt

1113

1114

1115

1116

1117

1118

1119

1120

1121

1122

1123

1124

1125

1126

1127

1128

You are a digital twin of a [NATIONALITY] voter. You are asked a policy issue or question. You must reason and then answer the question as if you were the voter. You reason and answer in English. Your final answer must be one of the answer options ("A" for "Yes", "B" for "No", "C" for "Neutral"). You must respond in the following format:  
 <reasoning>  
 Your reasoning goes here.  
 </reasoning>  
 <answer>  
 Final answer, one of  
 A) Yes,  
 B) No,  
 C) Neutral.  
 </answer>

1129

1130

where we replaced [NATIONALITY] with Swiss for *smartvote*, German for *Wahl-o-Mat*, and US for *ANES*.

1131

1132

1133

**LoRA settings.** We use the following LoRA settings for all models and methods. No hyperparameter optimization was performed.  $r = 32$ ,  $\alpha = 32$ , target modules: `q_proj`, `k_proj`, `v_proj`, `o_proj`, `gate_proj`, `up_proj`, `down_proj`.

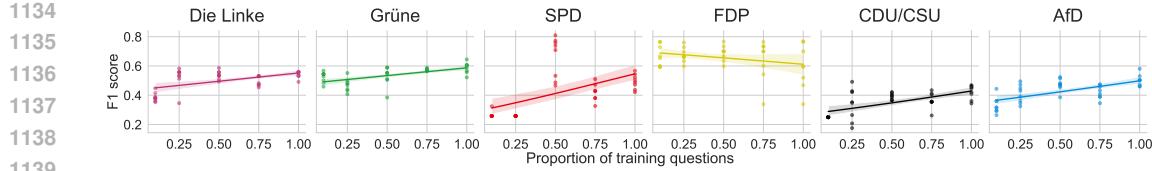


Figure 11: Impact of the number of questions in the training set. Shown are accuracies for 10%, 25%, 50%, 75%, and 100% of available training questions per party on the *WoM* dataset. Overall, F1 increases as the number of training questions increases. The largest increase is observed for the far-right AfD, while the FDP demonstrates a counterexample.

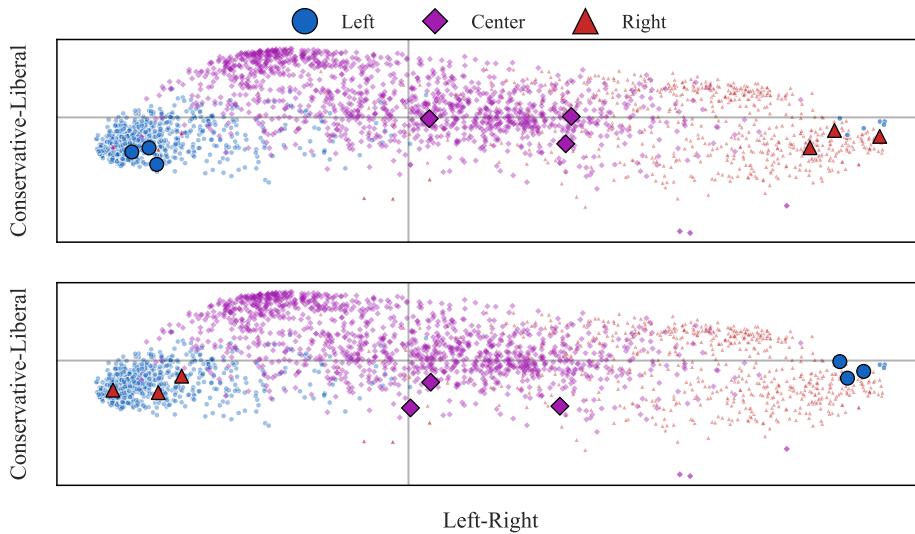


Figure 12: Inverting all answers in the dataset roughly corresponds to a rotation by  $180^\circ$  about the origin. Top: Candidates before inversion. Bottom: Candidates after inversion.

**ORPO.** We use the same hyperparameters as Stammbach et al. (2024) and the Hugging Face implementation of ORPO.

**SFT.** We train for 800 steps using the Adam optimizer with a batch size of 8, learning rate  $5e-5$ , a cosine learning rate scheduler, and we fix the maximum gradient norm at 1.0. The first 80 steps were used as a warm-up period. For reasoning-pretrained models, we change the `<reasoning>/</reasoning>` tags to their respective reasoning templates when fine-tuning. We used the Hugging Face implementation of SFT, and all other hyperparameters were left at the default values.

**GRPO.** We train for 800 steps using the Adam optimizer with a batch size of 8, learning rate  $5 \times 10^{-6}$ , a cosine learning rate scheduler, and we fix the maximum gradient norm at 1.0. We use the Adam optimizer with  $\beta_1 = 0.9$  and  $\beta_2 = 0.99$ , weight decay with 0.1. The first 80 steps were used as a warm-up period. The GRPO group size was 8. The GRPO  $\beta$  parameter was set to 0. The temperature was set at 1.0. For reasoning-pretrained models, we change the `<reasoning>/</reasoning>` tags to their respective reasoning templates in the format of reward functions. We used the Hugging Face implementation of GRPO, and all other hyperparameters were left at the default values.

**SFT+GRPO** . For this configuration, we used the same settings as for the individual SFT and GRPO configurations.

1188 **G CODE AND DATA ARTIFACTS**  
11891190 For full reproducibility, we fully release our code and data.  
11911192 • Code:  
1193     <https://anonymous.4open.science/r/Reasoning-Boosts-Opinion-Alignment-in-LLMs-7710/README.md>  
11941195 • Data:  
1196     <https://huggingface.co/datasets/anonymousemoji/Reasoning-Boosts-Opinion-Alignment-in-LLMs>  
1197  
1198  
1199  
1200  
1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208  
1209  
1210  
1211  
1212  
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221  
1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239  
1240  
1241