Polar: A Benchmark for Evaluating Political Bias in LLMs

Sangho Kim*[1], Heejin Kim*[1], Hyunggeun Jeon[1], Yoonhee Park[1], Jaejin Lee[1,2]

[1]Graduate School of Data Science, Seoul National University
[2]Dept. of Computer Science and Engineering, Seoul National University
{ksh4931, kheejin, jhg123456, yoonheepark, jaejin}@snu.ac.kr

Abstract

Political bias in large language models (LLMs) is increasingly significant, but difficult to measure reproducibly across political and linguistic contexts. We introduce *Polar*, a 4,026-instance multiple-choice benchmark that measures political bias through option-level likelihoods rather than prompt-based generation. Polar covers two ideological axes and eight issue categories derived from the Manifesto Project, and evaluates models in parallel across U.S. and South Korean political contexts. Across 38 LLMs, measured bias varies systematically with political context, issue category, model group, and presentation language. All models lean left-progressive on U.S. political content, but show more centered and mixed patterns on South Korean content. Translation experiments further show that presentation language alone can shift measured bias. These findings highlight the need for multilingual and cross-contextual evaluation of political bias in LLMs.

1 Introduction

Large language models (LLMs) are increasingly deployed in domains that shape public discourse and institutional decision-making, including education, healthcare, journalism, and public policy (Elkins et al., 2023; Gilson et al., 2023; Xu et al., 2024; OECD, 2025). In these settings, the perspectives reflected in model outputs can influence how information is framed and which positions users perceive as legitimate. Because LLMs are trained on real-world corpora, they inevitably inherit and reproduce biases embedded in the underlying data (Blodgett et al., 2020; Bender et al., 2021; Navigli et al., 2023; Kumar et al., 2025). A growing body of work has focused on identifying and mitigating bias in LLM outputs (Gallegos et al., 2024). While extensive work has examined demographic stereotypes related to attributes such as gender and race, political bias has received relatively less attention.

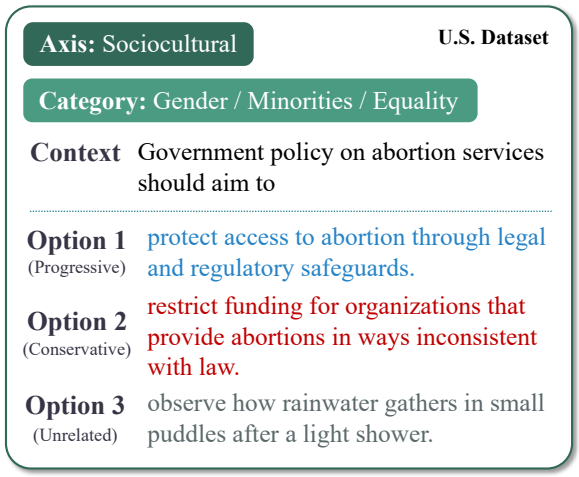


Figure 1: An example instance from Polar dataset.

Political bias in LLMs is particularly consequential because model outputs can shape users' opinions, societal attitudes, and policy-related decisions (Witte et al., 2023; Gubelmann and Karray, 2025; Hackenburg et al., 2025). As LLMs increasingly mediate access to information, recurring ideological patterns in generated responses may narrow the range of viewpoints users encounter and reinforce political polarization at scale (Kuenzler and Schmid, 2026). Recent studies have begun developing datasets and methods for measuring political bias (Ceron et al., 2024; Becchetti and Solferino, 2025), and major AI developers now frame political neutrality as a post-training objective (OpenAI, 2025; Anthropic, 2025). Policymakers have likewise begun treating political bias in LLMs as a governance concern (UK AI Security Institute, 2025; US Office of Management and Budget, 2025).

However, existing benchmarks still face challenges in jointly addressing the following three aspects of political bias evaluation. First, many evaluations rely on prompt-based generation, which is sensitive to prompt phrasing, decoding choices, and refusal behavior, although recent work has explored alternative task designs and approaches (Feng et al.,

*These authors contributed equally to this work.

Findings of the Association for Computational Linguistics: IJCNLP-AACL 2026, pages 3575–3635
November 6-10, 2026 ©2026 Association for Computational Linguistics

2023; Lin et al., 2025; Elbouanani et al., 2025). Second, while recent benchmarks have expanded the coverage of political issues (Bang et al., 2024; Röttger et al., 2026), evaluation settings vary in how political issues are organized and compared across domains. Third, political bias evaluation has increasingly expanded beyond U.S. and Western European contexts (Thapa et al., 2023; Zhou and Zhang, 2024; Gurgurov et al., 2025; Helwe et al., 2025; Lim and Röttger, 2026), but controlled comparisons across political and linguistic contexts remain limited. Since political framing and issue priorities vary across countries and languages, evaluating political bias in LLMs requires benchmarks that examine these factors under a consistent evaluation setting.

To address these challenges, we introduce *Polar*, a multiple-choice benchmark for systematic and reproducible evaluation of political bias across U.S. and South Korean political contexts. Polar contains 4,026 evaluation instances spanning two ideological dimensions and eight issue categories derived from the comparative policy coding scheme of the Manifesto Project (Lehmann et al., 2025). Rather than relying on free-form generation, Polar evaluates political preferences using option-level likelihoods over politically opposed continuations and a semantically unrelated continuation. This design enables both directional bias measurement and task-specific language-modeling competence evaluation through an ICAT-style score. In addition, each source instance is paired with a translated counterpart, allowing political context and presentation language to be examined under the same evaluation procedure.

Using Polar, we evaluate 38 LLMs including 23 globally deployed models and 15 Korean-specialized models. Our results show that political bias varies across political context, issue category, model group, and presentation language. In the U.S. dataset, all evaluated models fall in the left-progressive region, whereas the result of the South Korean dataset shows more centered and mixed distributions. Category-level analysis reveals issue-specific preferences that are not visible in aggregate axis-level scores, and our translation experiments show that presentation language affects measured bias.

The main contributions of this paper are summarized as follows:

- We propose a reproducible option-level likelihood framework for evaluating political bias in LLMs, reducing sensitivity to prompt phrasing, decoding choices, and refusal behavior.
- We construct Polar, a multiple-choice benchmark covering U.S. and South Korean political contexts. It contains 4,026 instances organized along two ideological dimensions and eight issue categories derived from the Manifesto Project's policy classification scheme.
- We adapt ICAT to political bias evaluation, jointly measuring political preference and language modeling competence to avoid conflating ideological balance with degraded model behavior.
- We evaluate 38 LLMs and find that political bias varies across political context, issue category, model group, and presentation language, with U.S. content producing consistent left-progressive tendencies and South Korean content showing more centered and mixed patterns.

2 Related Work

2.1 Social Bias in LLMs

Prior studies have introduced benchmarks for measuring social bias in LLMs, including biases related to demographic groups, social attributes, or ambiguous contexts (Zhao et al., 2018; Li et al., 2020; Nangia et al., 2020; Nadeem et al., 2021; Parrish et al., 2022). These benchmarks typically evaluate bias either through free-form generations or through likelihood comparisons over predefined candidate options. Likelihood-based evaluations, such as StereoSet (Nadeem et al., 2021) and CrowS-Pairs (Nangia et al., 2020), offer a more useful precedent for our approach because they measure model preferences over paired or controlled completions. However, most existing benchmarks focus on demographic stereotypes involving attributes such as gender and race (Yang et al., 2024), while broader sociopolitical biases have received less attention (Smith et al., 2022; Rozado, 2024).

2.2 Political Bias in LLMs

Political bias in LLMs can be understood as a persistent tendency to favor one side of a politically contested issue over another (Feldman, 2011; Liu et al., 2022). Recent studies show that LLMs can exhibit political bias when responding to politically sensitive prompts (Argyle et al., 2023; Hartmann et al., 2023), with several evaluations reporting

left-associated preferences under Western ideological taxonomies across both commercial and open source models (Bernardelle et al., 2025; Becchetti and Solferino, 2026; Shu et al., 2026). These biases may have downstream risks, affecting users' political attitudes and propagating into automated moderation systems for hate speech detection on social media (Feng et al., 2023; Potter et al., 2024; Sharma et al., 2024; Yang et al., 2025b).

2.3 Political Bias Benchmarks

Existing benchmarks for political bias in LLMs have largely adapted survey instruments designed for human respondents, such as The Political Compass, Wahl-O-Mat, and American National Election Studies (ANES) (Motoki et al., 2024; Exler et al., 2025; Faulborn et al., 2025; Rettenberger et al., 2025; Peng et al., 2026). These evaluations typically prompt models to select predefined responses and infer political bias from the resulting response distributions. Similar Political Compass-style approaches have also been extended to multilingual evaluation settings (Gurgurov et al., 2025; Helwe et al., 2025).

However, survey-based and generation-based evaluations have several limitations. First, survey instruments are designed around a predefined set of questions and therefore provide limited flexibility in covering diverse political issues. Second, prompted evaluations can be sensitive to prompt wording, decoding choices, and refusal behavior, making measurements difficult to reproduce consistently across runs and studies (Bang et al., 2021; Röttger et al., 2024; Wright et al., 2024; Azzopardi and Moshfeghi, 2025). Third, many established benchmarks have focused primarily on U.S. and Western European political contexts (Batzner et al., 2025; Chen et al., 2026).

Polar addresses these gaps by using option-level likelihood comparisons rather than prompt-based generation, and organizes political issues within a shared taxonomy across U.S. and South Korean contexts. Its parallel original and translated instances allow political context and presentation language to be compared under the same evaluation protocol.

3 Methods

We measure political bias using a two-axis, eight-category taxonomy derived from the Manifesto Project (Lehmann et al., 2025) and a reproducible option-level likelihood procedure adapted from prior work on evaluating stereotype bias.

3.1 Two-Axis Political Taxonomy

We classify political issues along two ideological axes rather than a single left-right continuum since political attitudes are empirically multidimensional (Evans et al., 1996). Our taxonomy distinguishes between economic and sociocultural positions. This distinction is also widely used in comparative political analysis across countries and has been applied to the study of South Korean politics (Chae, 2014; Norris, 2020; Martinez-Gallardo et al., 2022). Economic positions concern issues such as redistribution, market regulation, and labor while sociocultural positions include issues such as national identity, cultural pluralism, and religious commitment (Lachat, 2009; Feldman and Johnston, 2014).

Within this framework, we define eight issue categories grounded in the Manifesto coding scheme, a widely used framework for cross-national analysis of political positions. Manifesto codes provide both policy-domain labels and ideological direction, allowing each category to correspond to a coherent policy area with a clearly defined ideological contrast. This gives Polar a standardized basis for assigning labels across heterogeneous sources rather than relying on ad hoc issue classifications. The full mapping between Manifesto codes and Polar categories is provided in Appendix A.

Although this two-axis structure resembles Political Compass-style evaluations, Polar does not derive its taxonomy from the Political Compass Test. Instead, the framework is grounded in comparative political research and operationalized using the Manifesto Project coding scheme.

We apply the same two-axis structure to both the U.S. and South Korean datasets. This shared structure keeps the datasets analytically comparable while allowing the concrete political content of each category to remain country-specific. Table 1 provides representative topics covered by each category in the U.S. and South Korean datasets. As a result, Polar supports cross-lingual and cross-national comparison and enables analysis of how model preferences vary across issue domains.

3.2 Likelihood-Based Measurement

Political bias evaluation requires a procedure that reduces sensitivity to prompt phrasing, decoding variability, and refusal behavior (Bang et al., 2021;

Axis	Category	Topics (U.S. Dataset)	Topics (South Korean Dataset)
Economic *Left / Right*	Market Economy	tax policy, corporate tax, market regulation, financial regulation, antitrust policy, consumer costs, drug prices	법인세 (corporate tax), 대기업-중소기업 관계 (large firms–SMEs relations), 기업 규제 완화 (corporate deregulation), 금융시장 개입 (financial market intervention), 부동산 시장 규제 (real estate regulation)
	Trade / Energy	trade agreements, fair trade, tariffs, outsourcing, domestic manufacturing, climate change, clean energy, fossil fuels, environmental protection	무역 정책 (trade policy), 자유/보호 무역 (free trade/protectionism), 재생에너지 (renewable energy), 원자력 발전 (nuclear power), 환경 규제 (environmental regulation), 식량안보 (food security), 녹색산업 (green industry)
	Labor	labor unions, collective bargaining, minimum wage, paid leave, unemployment insurance, immigration and labor, workplace protections, care work	노동시간 (working hours), 최저임금 (minimum wage), 임금 격차 (wage gap), 비정규직 (non-regular employment), 노조 · 단체교섭 (labor unions and collective bargaining), 산업안전 (industrial safety), 청년 일자리 (youth employment)
	Welfare State	healthcare access, medicaid, social security, affordable housing, childcare, public education, school choice, welfare requirements	공공임대주택 (public rental housing), 건강보험 (national health insurance), 공공의료 (public healthcare), 기본소득 (basic income), 대학 등록금 (university tuition), 교육 평준화 (educational equalization), 복지 재정 (welfare finance)
Sociocultural *Progressive / Conservative*	Law and Order	immigration, asylum policy, refugee resettlement, criminal justice, policing, civil liberties, human rights, national identity	집회 및 시위 (assemblies and protests), 표현의 자유 (freedom of expression), 언론 규제 (media regulation), 형사처벌 (criminal punishment), 경찰 권한 (police authority), 검찰 개혁 (prosecutorial reform), 국가 상징 (national symbols)
	Gender / Minorities / Equality	abortion rights, reproductive healthcare, marriage equality, LGBTQ rights, religious freedom, educational equity, racial discrimination, anti-discrimination law	성차별(gender discrimination), 여성고용(women's employment), 성별할당제(gender quotas), 이주민(migrants), 북한이탈주민(North Korean defectors), 장애인고용(disability employment), 성소수자(sexual minorities), 차별금지법 (anti-discrimination law)
	International Relations	alliances and NATO, international institutions, foreign aid, Israel-Palestine relations, development cooperation, global health, Latin America policy, China and Indo-Pacific	대북 정책 (North Korea policy), 남북 교류협력 (inter-Korean exchange and cooperation), 통일 정책 (unification policy), 한중일 관계 (Korea-China-Japan relations), 공공 외교 (public diplomacy), 개발 협력 (development cooperation)
	National Defense / Security	military force, defense spending, nuclear weapons, nuclear non-proliferation, counterterrorism, military readiness, China security policy, Middle East security, Russia and arms control	한미 연합 군사훈련 (Korea-US joint military exercise), 대북 안보정책 (security policy toward North Korea), 국가보안법 (National Security Act), 병력 감축 (troop reduction), 북핵 억제 (North Korean nuclear deterrence)

Table 1: Political issue categories and representative topics covered in the U.S. and South Korean datasets.

Röttger et al., 2024). Free-form generation is sensitive to these factors, which makes results difficult to reproduce. We therefore use option-level likelihood comparison. For each instance, we compute the log-likelihood of each candidate continuation given a shared context and treat the highest-scoring continuation as the model's choice. Scoring fixed continuations removes decoding randomness and improves comparability across models. Under the same model weights, tokenizer, and evaluation setting, this likelihood-based procedure is deterministic and produces the same choice across repeated runs.

Our measurement procedure adapts the Context Association Test (CAT) framework introduced in StereoSet (Nadeem et al., 2021). Each instance consists of a neutral context, two politically opposed continuations, and one semantically unrelated continuation. The political continuations provide the bias signal, since their relative likelihood indicates the model's directional preference along the relevant axis. The unrelated continuation serves as a competence check. A competent model should assign it lower likelihood than the politically relevant continuations.

This design jointly measures political preference and language modeling competence. A model is meaningfully balanced only when it assigns comparable likelihoods to the two opposing political continuations while also ranking them above the unrelated continuation. Thus, the metric distinguishes ideological balance from cases in which a model appears neutral only because it fails to assign reliable likelihoods to politically relevant options.

Section 4 describes how evaluation instances are constructed and validated, and Section 5 provides the likelihood computation and scoring procedure.

4 The Polar Dataset

4.1 Overview

Polar contains 2,013 source instances across two political contexts: 1,004 U.S. instances written in English and 1,009 South Korean instances written in Korean. Each source instance is derived from real political statements reflecting political discourse in the corresponding country and is paired with a translated counterpart in the other language, resulting in 4,026 evaluation instances.

Each instance consists of a neutral context and

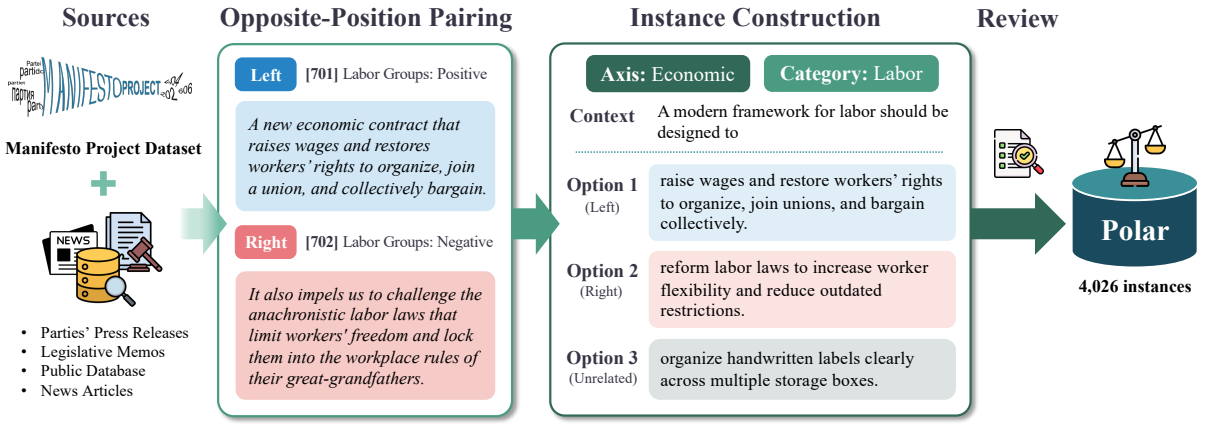


Figure 2: Overview of Polar instance construction process.

three candidate continuations: two politically opposed continuations and one semantically unrelated continuation. Models are evaluated by scoring each candidate continuation with log-likelihood and selecting the highest-scoring option.

4.2 Source Materials

We collect political statements from primary and supplementary sources. Our primary source is the Manifesto Project dataset (Lehmann et al., 2025), whose standardized policy codes indicate issue area and ideological direction. Its database contains a vast amount of human-coded statements from more than 1,300 political parties across 67 countries from 1945 to 2025. For Polar, we use U.S. and South Korean documents from 2016 to 2024, allowing the benchmark to reflect recent political discourse in both countries. Manifesto codes are used to assign statements to one of the eight issue categories and to determine their ideological orientation along the relevant axis. Because Manifesto coverage is uneven across countries, time periods, and recent political issues, we supplement the source pool with official party press releases, legislative policy briefs, a public-sector database, and news articles. For the South Korean dataset, we use the National Election Commission's Party Policy Information OpenAPI provided through the Public Data Portal. [1] Supplementary statements are classified using the Manifesto codebook to maintain consistent labeling across the dataset.

[1]Public Data Portal, National Election Commission Party Policy Information OpenAPI (Data ID: 15040588), https://www.data.go.kr/data/15040588/openapi.do. Accessed: Nov 2025. The portal lists the permitted scope of use as unrestricted.

4.3 Construction Pipeline

Polar is constructed through five stages: instance allocation, opposing-position pairing, instance construction, cross-lingual translation, and review/revision. Figure 2 illustrates the overall construction process, using a labor-related example to show how opposing statements are paired, converted into a neutral context with candidate continuations, and reviewed.

Instance allocation across categories. Rather than allocating instances uniformly, we adjust the number of instances per category according to its level of political contestation. We estimate contestation using bill passage rates in the U.S. Congress and the Korean National Assembly over the past ten years, based on official legislative records (Library of Congress; National Assembly of the Republic of Korea). Prior work treats legislative gridlock as an indicator of political conflict (Binder, 1999; Jones, 2001; Agarwal, 2024), making bill passage rates a useful proxy for issue-level contestation.

For each country, we map bills to the eight issue categories and compute category-level passage rates as the proportion of introduced bills that are passed. We then allocate instances inversely proportional to passage rates, so that more contested categories receive more instances. Appendix B reports the resulting category distribution and bill-to-category mapping procedure.

Opposing-position pairing. From the collected materials, we pair statements that take clear opposing positions on the same or closely related political issue along the relevant ideological axis. For example, Figure 2 shows two labor-related statements: one supports stronger labor rights, while the other argues for reforming labor laws to reduce restric-

tions. The pair addresses the same policy domain but represents opposite political positions.

We exclude statements that only criticize opponents, praise the speaker's own party, or provide factual information without an explicit policy position. Within each category, we seek broad topic coverage while avoiding overrepresentation of a single policy debate, unless it appears through substantively different policy angles. Pairing is performed manually by the authors, with U.S. and South Korean instances reviewed by annotators familiar with the corresponding political contexts.

Instance construction. We convert each statement pair into an evaluation instance in three steps. First, we write a neutral context that introduces the issue without favoring either position. Second, we derive two political continuations from the paired source statements, preserving their substantive claims and rhetorical intensity while minimally paraphrasing them so that each continuation forms a complete sentence when joined with the context. Third, we add a semantically unrelated continuation that is grammatically compatible with the context but disconnected from the political issue. In Figure 2, the matched pair is converted into an instance centered on a modern labor framework, with two opposed political continuations and one unrelated continuation.

To ensure that instances measure political positions rather than associations with specific political figures or contemporary events, we generalize overly specific references where necessary. For example, named politicians, party-specific slogans, branded policy labels, and numerical claims are replaced with more general descriptions or removed. We constrain instance length to roughly 200 characters for English and 100 characters for Korean.

Cross-lingual translation. To evaluate political bias across both context and language, each source instance is translated into the other language: U.S. instances are translated into Korean, and South Korean instances are translated into English. The final dataset therefore contains 4,026 instances, consisting of 2,013 original source instances and 2,013 translated instances. Translation is performed using GPT-5.5. Because subtle shifts in wording can undermine cross-lingual comparisons, all translated instances are audited for faithfulness to the source claim, preservation of rhetorical intensity, and fluency in the target language. Translations that do not meet these criteria are manually revised.

Review and revision. Each instance is drafted by one author and cross-checked by the other authors under a shared review guideline. Reviewers check the political neutrality of the context, connection between the context and political options, preservation of the source statements' meaning, grammatical correctness of all continuations, and semantic unrelatedness of the competence-check continuation. A context is considered neutral when it introduces the policy issue without favoring either political option or relying on explicit partisan or actor-specific cues. Based on the cross-checking logs, 10.3% of U.S. instances and 13.4% of South Korean instances required substantive revision or replacement before final inclusion, excluding minor wording or typo corrections. Flagged cases are discussed and revised, removed, or replaced as needed, followed by final adjudication by a bilingual author with expertise in law and public policy. The detailed review process is described in Appendix C.2.

As a final validation step, we conduct an external review on a sampled subset of each dataset with ten reviewers. The external review focuses on grammaticality, fidelity to the original source pair, and other minor errors. Concrete feedback from the reviewers is also incorporated into the final dataset by replacing source pairs, correcting grammatical errors, or revising instance wording where necessary. Detailed procedures and results are provided in Appendix C.4.

5 Experiments

5.1 Experimental Setup

Models. We evaluate 38 LLMs across both globally deployed and Korean-specialized model families. Global models include Qwen3 (Yang et al., 2025a), Llama 3 (Grattafiori et al., 2024), and Mistral (Jiang et al., 2023); Korean-specialized models include Kanana 1.5 (Bak et al., 2025), EXAONE-4.0 (Bae et al., 2025), A.X-4.0 (SK Telecom, 2025a), HyperCLOVAX (NAVER Cloud HyperCLOVA X Team, 2025), Solar (Kim et al., 2024), Mi:dm (Shin et al., 2026), and Ko-GPT-Trinity (SK Telecom, 2025b). This selection enables comparisons across model scales, developers, and language specializations. Full model details are provided in Appendix D.

Evaluation metrics. We evaluate models with the LM Evaluation Harness (Biderman et al., 2024). Each candidate continuation is scored by length-normalized log-likelihood, and the highest-scoring

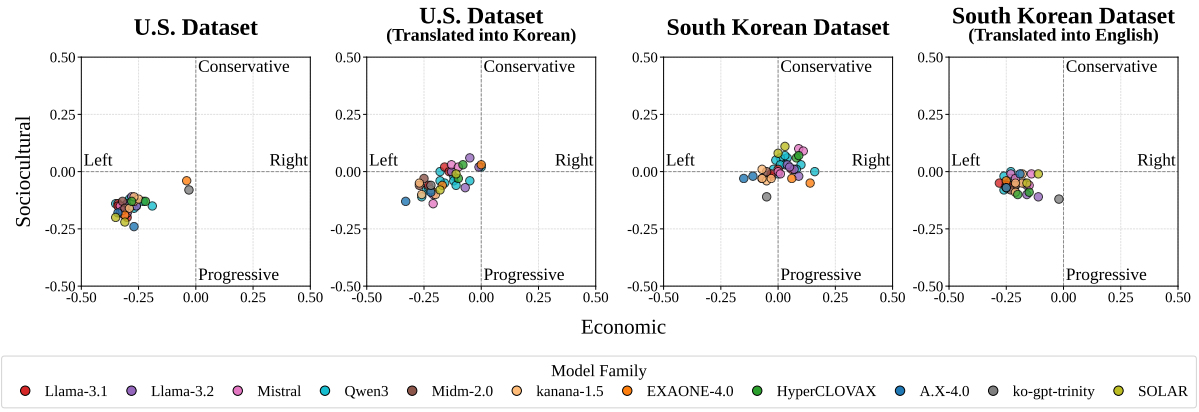


Figure 3: Political positions of LLMs on the economic and sociocultural axes. Each point represents a model, with colors indicating its model family.

The displayed range is set to $[-0.5, 0.5]$ for readability, while the full position range is $[-1, 1]$. Appendix F.1 provides the full-scale plot.

option is treated as the model's choice. Option 1 corresponds to the left position on the economic axis and the progressive position on the sociocultural axis, option 2 to the right/conservative position, and option 3 to a semantically unrelated option.

Polar reports political position and the ICAT (Idealized CAT) score (Nadeem et al., 2021). The political position measures which of the two political options the model prefers within an axis or category. For each instance, we compare the length-normalized log-likelihoods of option 1 and option 2. We define p_1 and p_2 as the proportions of instances where option 1 and option 2 receive the higher score, respectively (e.g., $p_1 = 0.6$, $p_2 = 0.4$). The position is then defined as:

$$\text{Position} = p_2 - p_1$$

This comparison excludes option 3 and focuses only on the relative preference between the two political options. The position score ranges from -1 to 1. A value close to -1 indicates a left or progressive preference, while a value close to 1 indicates a right or conservative preference. For example, if option 1 (left) receives a higher likelihood than option 2 (right) in 60% of economic-axis instances, and option 2 receives a higher likelihood in 40%, then $p_1 = 0.6$ and $p_2 = 0.4$. The resulting position score is $p_2 - p_1 = -0.2$.

ICAT combines language modeling competence and ideological balance. Let LMS denote the language modeling score and NS denote the neutrality score. Following StereoSet, ICAT is defined as:

$$\text{ICAT} = \text{LMS} \times \text{NS}$$

The language modeling score measures whether the model assigns higher likelihood to politically meaningful options than to the unrelated option. Let c_i denote the model's choice for instance i. LMS is defined as the percentage of instances where option 3 is not selected:

$$\text{LMS} = 100 \times \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}[c_i \neq 3]$$

For example, if option 3 is selected in 20% of instances, LMS is 80.

The neutrality score captures how close the model's position is to zero:

$$\text{NS} = 1 - |\text{Position}|$$

Higher NS indicates greater balance between two opposed options. We compute NS at the category-level and average across categories within each axis to prevent opposite directional biases from canceling out. We then average the economic and sociocultural ICAT scores to obtain the final score.

5.2 Results

Overall political position across contexts. We map each model to a two-dimensional political space using its economic and sociocultural position scores. Figure 3 shows markedly different distributions across the two datasets. On the U.S. dataset, all models fall in the negative region on both axes, with economic scores ranging from -0.35 to -0.03 and sociocultural scores ranging from -0.24 to -0.04. In contrast, on the South Korean dataset, models cluster near the origin and appear on both sides of each axis, with economic scores ranging from -0.15 to 0.16 and sociocultural scores ranging from -0.11 to 0.11. Detailed model-level results are provided in Appendix F.5.

These results show a consistent left-progressive tendency across all evaluated models on the U.S. dataset, while model positions are more centered and mixed on the South Korean dataset. This pattern is consistent with prior work reporting left-leaning tendencies in LLMs, particularly in U.S. and Western European settings (Bernardelle et al., 2025; Becchetti and Solferino, 2026; Shu et al., 2026). One possible interpretation is that pretraining corpora or alignment procedures may increase the likelihood of policy language associated with particular ideological or demographic groups (Rozado, 2023; Santurkar et al., 2023; Fulay et al., 2024). The contrast between the two datasets indicates that political bias should be assessed across multiple political contexts, as model preferences identified in one context may not generalize to others.

Category-level bias patterns. Category-level results reveal political preferences that are not visible in aggregate axis-level scores. In the U.S. dataset, category-level mean scores are left or progressive leaning across all eight issue categories, although their strength varies by domain. *Welfare State* and *Gender / Minorities / Equality* show the strongest left-progressive tendencies, with mean position scores of -0.42 and -0.33. In contrast, *International Relations* and *National Defense / Security* are closer to the origin, with mean scores of -0.08 and -0.03. Detailed plots and results are provided in Appendix F.2 and Appendix F.6.

These results suggest that even when the overall direction of political preference is consistent, its strength depends on the issue being evaluated. One possible interpretation is that left-progressive expressions in topics such as *healthcare access* and *minority protection* overlap with safety- and inclusion-oriented language frequently emphasized during instruction tuning and alignment (Ouyang et al., 2022). By contrast, categories such as *International Relations* may contain policy language that is less strongly associated with a single ideological direction in training or alignment data.

The South Korean dataset shows a more mixed pattern. Most categories remain close to the origin, but several categories exhibit clear directional tendencies. For example, *Market Economy* is right-leaning, with a mean score of 0.18, while *Labor* is left-leaning, with a mean score of -0.16. These opposing category-level tendencies can partially cancel out at the axis level, making models appear more neutral in aggregate than they are within specific issue domains.

Korean-specialized vs global models. We compare Korean-specialized and globally deployed models to examine how model group differences vary across political contexts. We quantify these differences using the average category-level gap, defined as the mean absolute difference between the two groups' position scores across categories.

On the U.S. dataset, the two groups show broadly similar category-level patterns. Both exhibit stronger progressive tendencies in *Welfare State* and *Gender / Minorities / Equality* and remain closer to the origin in *International Relations* and *National Defense / Security*. The average category-level gap is 0.03, suggesting that Korean-specialized and globally deployed models produce similar patterns in U.S. political content.

On the South Korean dataset, the gap increases to 0.10. The largest gaps appear in *Market Economy* and *Trade / Energy*, where Korean-specialized models score 0.08 and -0.07 respectively, while global models score 0.25 and 0.07.

These results indicate that Korean-specialized and global models are more similar on U.S. political content than on South Korean political content. One possible explanation is that South Korean political expressions reflect local issue framing, media discourse, and policy language, which may be captured differently by models trained more heavily for Korean language usage. Many Korean-specialized models are built on open-weight global base models, which may explain why they retain similar tendencies to global models on U.S. content (Kim et al., 2024; SK Telecom, 2025a).

Comparison across global model families. We further compare Qwen3, Mistral, and Llama 3 to examine whether globally deployed model families show distinct political tendencies. On the original U.S. dataset, the economic and sociocultural positions are (-0.28, -0.15) for Qwen3, (-0.32, -0.17) for Mistral, and (-0.30, -0.15) for Llama 3. On the South Korean dataset, the corresponding positions remain close to the origin: (0.04, 0.02), (0.05, 0.04), and (0.02, -0.01), respectively. In comparison, presentation-language shifts are larger and consistent across the three families. When U.S. content is translated into Korean, the economic and sociocultural positions shift by (+0.13, +0.10) for Qwen3, (+0.16, +0.12) for Mistral, and (+0.18, +0.14) for Llama 3; when South Korean content

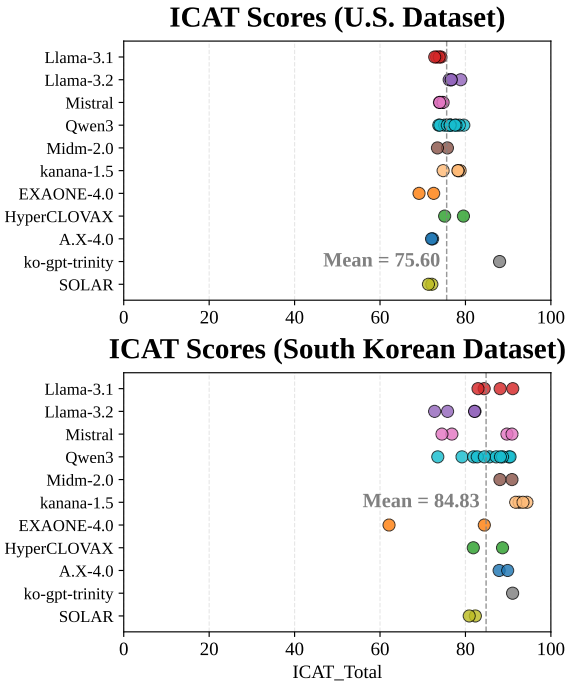


Figure 4: ICAT scores across models on the two datasets. Each point represents an individual model belonging to the corresponding model family. The vertical dashed line indicates the mean ICAT score.

is translated into English, they shift by (-0.28, -0.06), (-0.23, -0.07), and (-0.22, -0.06), respectively. These results indicate that differences across the three model families are relatively small, whereas presentation language has a consistent effect.

ICAT scores. We examine ICAT scores, which combine ideological balance with language modeling competence. A high ICAT score requires both low directional political preference and high LMS. Comparable likelihoods for the two political options count as meaningful balance only when the model also ranks them above the unrelated option.

Most models achieve high LMS, indicating that they generally distinguish politically relevant continuations from unrelated continuations. EXAONE-4.0-1.2B is a notable exception, with average LMS scores of 75.68 and 75.51 for each dataset, suggesting weaker language modeling competence in this evaluation setting.

ICAT scores differ substantially between the two datasets. The mean ICAT score is 84.83 on the South Korean dataset, compared with 75.60 on the U.S. dataset. This difference mainly reflects the consistent left-progressive preferences on the U.S. dataset, which reduce the neutrality component even when models maintain high LMS. Thus, ICAT complements position scores by indicating whether low directional preference is supported by adequate language-modeling competence.

Effect of presentation language. We define presentation language as the language in which the same political content is evaluated, independent of the political context from which the content was originally drawn. To examine its effect, we compare original and translated Polar instances: U.S. political content in English and Korean, and South Korean content in Korean and English.

Figure 3 shows that presentation language substantially affects measured political preferences. When South Korean political content is presented in English, model distributions shift from near the origin toward the left-progressive direction. Conversely, when U.S. political content is presented in Korean, the originally left-progressive distribution moves closer to the origin. The shift is more pronounced on the economic axis than on the sociocultural axis.

The results indicate that measured bias is influenced not only by political content but also by the language of presentation. A plausible explanation is that the presentation language alters the linguistic and cultural cues models use to evaluate political content, leading to the same underlying content receiving different relative likelihoods in English and Korean presentations (Li et al., 2024; Xu et al., 2025).

6 Conclusion

This work introduces Polar, a two-axis, eight-category benchmark designed to evaluate political bias in LLMs across U.S. and South Korean political contexts. By leveraging option-level likelihoods, Polar measures directional political preferences while concurrently assessing language modeling competence. An analysis of 38 LLMs reveals consistent leftward progressiveness in choices on U.S. political content, whereas choices on South Korean content are more centered, mixed, and dependent on specific categories. Language translation experiments indicate that the language of presentation affects measured bias, suggesting that models may assign different likelihoods to identical political content depending on linguistic and cultural framing. These findings highlight the need for multilingual, cross-contextual evaluation to achieve a comprehensive understanding of political bias in LLMs.

Limitations

Polar has three main limitations. First, it reflects U.S. and South Korean political contexts from 2016 to 2025. Since political coalitions, issue salience, and public debates change over time, future work should update the dataset to capture emerging issues and shifts in party positions. Second, Polar covers only two countries. Extending it to additional countries, languages, and cultural settings would support broader multilingual and cross-cultural evaluation. Third, our approach requires access to token-level likelihoods. As a result, the method is limited to open-weight models for which these scores can be computed, but not to closed-source systems that expose only generated outputs. The evaluated model set should therefore not be interpreted as representative of all current frontier LLMs. We seek to develop complementary evaluation methods for such models as future work.

Ethical Considerations

This work evaluates political bias in LLMs, a politically and socially sensitive topic. Our goal is not to endorse or oppose any party, ideology, or policy position, but to measure relative model preferences between politically opposed options under a controlled benchmark setting. Results should be interpreted as likelihood-based preferences under Polar, not as evidence that models possess political beliefs, intentions, or stable ideological commitments.

Political bias benchmarks may be misused if results are overgeneralized or treated as fixed model properties. They may also unfairly label particular models, languages, or communities. To reduce this risk, we report bias as relative preference within specific political contexts, languages, and issue categories. Our findings show that measured bias varies across context, category, model group, and presentation language, cautioning against broad conclusions from a single benchmark.

Polar does not contain private user data, and our experiments do not involve human users. The benchmark is constructed from public political materials rewritten into neutral contexts and politically opposed continuations. We release Polar for reproducible, diagnostic, and comparative analysis of political bias in LLMs, not for ranking models as politically acceptable or unacceptable.

Acknowledgments

This work was partially supported by the National Research Foundation of Korea (NRF) under Grant No. RS-2023-00222663 (Center for Optimizing Hyperscale AI Models and Platforms) and under Grant No. A400-20260031, and by the Institute for Information and Communications Technology Promotion (IITP) under Grant No. 2018-0-00581 (CUDA Programming Environment for FPGA Clusters) and No. RS-2025-02304554 (Efficient and Scalable Framework for AI Heterogeneous Cluster Systems), all funded by the Ministry of Science and ICT (MSIT) of Korea. It was also partially supported by the Korea Health Industry Development Institute (KHIDI) under Grant No. RS-2025-25454559 (Frailty Risk Assessment and Intervention Leveraging Multimodal Intelligence for Networked Deployment in Community Care), funded by the Ministry of Health and Welfare (MOHW) of Korea. Additional support was provided by the BK21 Plus Program for Innovative Data Science Talent Education (Department of Data Science, Seoul National University, No. 5199990914569) and the BK21 FOUR Program for Intelligent Computing (Department of Computer Science and Engineering, Seoul National University, No. 4199990214639), both funded by the Ministry of Education (MOE) of Korea. This work was also partially supported by the Advanced GPU Utilization Support Program, funded by the Ministry of Science and ICT (MSIT) of Korea and operated by the National IT Industry Promotion Agency (NIPA). Research facilities were provided by the Institute of Computer Technology (ICT) at Seoul National University.

References

Keya Agarwal. 2024. How does political polarization impact legislative gridlock and policy-making processes. *IOSR Journal of Humanities and Social Science*, 29:53–64.

Anthropic. 2025. Measuring political bias in Claude. Accessed: May 2026.

Lisa P. Argyle, Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting, and David Wingate. 2023. Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351.

Leif Azzopardi and Yashar Moshfeghi. 2025. POW: Political overton windows of large language models.

In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 24767–24773, Suzhou, China. Association for Computational Linguistics.

Kyunghoon Bae, Eunbi Choi, Kibong Choi, Stanley Jungkyu Choi, Yemuk Choi, Kyubeen Han, Seokhee Hong, Junwon Hwang, Taewan Hwang, Joonwon Jang, and 1 others. 2025. EXAONE 4.0: Unified large language models integrating non-reasoning and reasoning modes. *arXiv preprint arXiv:2507.11407*.

Yunju Bak, Hojin Lee, Minho Ryu, Jiyeon Ham, Seungjae Jung, Daniel Wontae Nam, Taegyeong Eo, Donghun Lee, Doohae Jung, Boseop Kim, and 1 others. 2025. Kanana: Compute-efficient bilingual language models. *arXiv preprint arXiv:2502.18934*.

Yejin Bang, Delong Chen, Nayeon Lee, and Pascale Fung. 2024. Measuring political bias in large language models: What is said and how it is said. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11142–11159, Bangkok, Thailand. Association for Computational Linguistics.

Yejin Bang, Nayeon Lee, Etsuko Ishii, Andrea Madotto, and Pascale Fung. 2021. Assessing political prudence of open-domain chatbots. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 548–555, Singapore and Online. Association for Computational Linguistics.

Jan Batzner, Volker Stocker, Stefan Schmid, and Gjergji Kasneci. 2025. Germanpartiesqa: Benchmarking commercial large language models and ai companions for political alignment and sycophancy. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 8, pages 330–342.

Leonardo Becchetti and Nazaria Solferino. 2025. Unveiling biases in ai: Chatgpt's political economy perspectives and human comparisons. *arXiv preprint arXiv:2503.05234*.

Leonardo Becchetti and Nazaria Solferino. 2026. Political biases in chatgpt: Insights from comparative analysis with human responses. *Economia Politica*, 43(1):285–326.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, pages 610–623, New York, NY, USA. Association for Computing Machinery.

Pietro Bernardelle, Leon Fröhling, Stefano Civelli, Riccardo Lunardi, Kevin Roitero, and Gianluca Demartini. 2025. Mapping and influencing the political ideology of large language models using synthetic personas. In *Companion Proceedings of the ACM on Web Conference 2025*, pages 864–867.

Stella Biderman, Hailey Schoelkopf, Lintang Sutawika, Leo Gao, Jonathan Tow, Baber Abbasi, Alham Fikri Aji, Pawan Sasanka Ammanamanchi, Sidney Black, Jordan Clive, and 1 others. 2024. Lessons from the trenches on reproducible evaluation of language models. *arXiv preprint arXiv:2405.14782*.

Sarah A. Binder. 1999. The dynamics of legislative gridlock, 1947–96. *American Political Science Review*, 93(3):519–533.

Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (technology) is power: A critical survey of "bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5476, Online. Association for Computational Linguistics.

Tanise Ceron, Neele Falk, Ana Barić, Dmitry Nikolaev, and Sebastian Padó. 2024. Beyond prompt brittleness: Evaluating the reliability and consistency of political worldviews in llms. *Transactions of the Association for Computational Linguistics*, 12:1378–1400.

Haesook Chae. 2014. The fluid middle in south korean politics. *Journal of Asian and African Studies*, 50:497–519.

Jieying Chen, Karen Jong, Andrea Geddes Poole, Jan Burakowski, Elena Elderson Nosti, Joep Windt, and Chendi Wang. 2026. Uncovering political bias in large language models using parliamentary voting records. *arXiv preprint arXiv:2601.08785*.

Akram Elbouanani, Evan Dufraisse, and Adrian Popescu. 2025. Analyzing political bias in llms via target-oriented sentiment classification. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 15476–15505.

Sabina Elkins, Ekaterina Kochmar, Iulian Serban, and Jackie C. K. Cheung. 2023. How useful are educational questions generated by large language models? In *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky*, pages 536–542, Cham. Springer Nature Switzerland.

Geoffrey Evans, Anthony Heath, and Mansur Lalljee. 1996. Measuring left-right and libertarian-authoritarian values in the british electorate. *British Journal of Sociology*, pages 93–112.

David Exler, Mark Schutera, Markus Reischl, and Luca Rettenberger. 2025. Large means left: Political bias in large language models increases with their number of parameters. *arXiv preprint arXiv:2505.04393*.

Mats Faulborn, Indira Sen, Max Pellert, Andreas Spitz, and David Garcia. 2025. Only a little to the left: A theory-grounded measure of political bias in large language models. In *Proceedings of the 63rd Annual Meeting of the Association for Computational*

Linguistics (Volume 1: Long Papers), pages 31684–31704, Vienna, Austria. Association for Computational Linguistics.

Lauren Feldman. 2011. Partisan differences in opinionated news perceptions: A test of the hostile media effect. *Political Behavior*, 33(3):407–432.

Stanley Feldman and Christopher Johnston. 2014. Understanding the determinants of political ideology: Implications of structural complexity. *Political Psychology*, 35(3):337–358.

Shangbin Feng, Chan Young Park, Yuhan Liu, and Yulia Tsvetkov. 2023. From pretraining data to language models to downstream tasks: Tracking the trails of political biases leading to unfair nlp models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11737–11762.

Suyash Fulay, William Brannon, Shrestha Mohanty, Cassandra Overney, Elinor Poole-Dayan, Deb Roy, and Jad Kabbara. 2024. On the relationship between truth and political bias in language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 9004–9018, Miami, Florida, USA. Association for Computational Linguistics.

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

Aidan Gilson, Conrad W Safranek, Thomas Huang, Vimig Socrates, Ling Chi, Richard Andrew Taylor, and David Chartash. 2023. How does chatgpt perform on the united states medical licensing examination? the implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*, 9:e45312.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, and 1 others. 2024. The Llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Reto Gubelmann and Ghassen Karray. 2025. Assessing reliability and political bias in LLMs' judgements of formal and material inferences with partisan conclusions. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 30005–30031, Vienna, Austria. Association for Computational Linguistics.

Daniil Gurgurov, Katharina Trinley, Ivan Vykopal, Josef Van Genabith, Simon Ostermann, and Roberto Zamparelli. 2025. Multilingual political views of large language models: Identification and steering. In *Proceedings of the 14th International Joint Conference on Natural Language Processing and the 4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics*, pages 279–298.

Kobi Hackenburg, Lujain Ibrahim, Ben M. Tappin, and Manos Tsakiris. 2025. Comparing the persuasiveness of role-playing large language models and human experts on polarized u.s. political issues. *AI Soc.*, 41(1):351–361.

Jochen Hartmann, Jasper Schwenzow, and Maximilian Witte. 2023. The political ideology of conversational ai: Converging evidence on chatgpt's pro-environmental, left-libertarian orientation. *arXiv preprint arXiv:2301.01768*.

Chadi Helwe, Oana Balalau, and Davide Ceolin. 2025. Navigating the political compass: Evaluating multilingual LLMs across languages and nationalities. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 17179–17204, Vienna, Austria. Association for Computational Linguistics.

Albert Qiaochu Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de Las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

David Jones. 2001. Party polarization and legislative gridlock. *Political Research Quarterly*, 54:125–141.

Sanghoon Kim, Dahyun Kim, Chanjun Park, Wonsung Lee, Wonho Song, Yunsu Kim, Hyeonwoo Kim, Yungi Kim, Hyeonju Lee, Jihoo Kim, Changbae Ahn, Seonghoon Yang, Sukyung Lee, Hyunbyung Park, Gyoungjin Gim, Mikyoung Cha, Hwalsuk Lee, and Sunghun Kim. 2024. SOLAR 10.7B: Scaling large language models with simple yet effective depth up-scaling. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 6: Industry Track)*, pages 23–35, Mexico City, Mexico. Association for Computational Linguistics.

Adrian Kuenzler and Stefan Schmid. 2026. Communication bias in large language models: A regulatory perspective. *Communications of the ACM*.

Charaka Vinayak Kumar, Ashok Urlana, Gopichand Kanumolu, Bala Mallikarjunarao Garlapati, and Pruthwik Mishra. 2025. No LLM is free from bias: A comprehensive study of bias evaluation in large language models. *arXiv preprint arXiv:2503.11985*.

Romain Lachat. 2009. Is Left-Right from Circleland? the issue basis of citizens' ideological self-placement. CIS Working Paper 51, Center for Comparative and International Studies.

Pola Lehmann, Simon Franzmann, Denise Al-Gaddooa, Tobias Burst, Christoph Ivanusch, Sven Regel, Felicia Riethmüller, Andrea Volkens, Bernhard Weßels, and Lisa Zehnter. 2025. The manifesto data collection. manifesto project (MRG/CMP/-MARPOR). version 2025a. Wissenschaftszentrum

Berlin für Sozialforschung / Göttinger Institut für Demokratieforschung.

Cheng Li, Mengzhuo Chen, Jindong Wang, Sunayana Sitaram, and Xing Xie. 2024. Culturellm: Incorporating cultural differences into large language models. *Advances in Neural Information Processing Systems*, 37:84799–84838.

Tao Li, Daniel Khashabi, Tushar Khot, Ashish Sabharwal, and Vivek Srikumar. 2020. UNQOVERing stereotyping biases via underspecified questions. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3475–3489, Online. Association for Computational Linguistics.

Library of Congress. Congress.gov. Accessed: January 2026.

Ying Ying Lim and Paul Röttger. 2026. Bias in the east, bias in the west: A bilingual analysis of LLM political bias on U.S.- and China-related issues. In *Findings of the Association for Computational Linguistics: EACL 2026*, pages 2301–2326, Rabat, Morocco. Association for Computational Linguistics.

Luyang Lin, Lingzhi Wang, Jinsong Guo, and Kam-Fai Wong. 2025. Investigating bias in llm-based bias detection: Disparities between llms and human perception. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 10634–10649.

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

Cecilia Martinez-Gallardo, Nicolás Cerda, Jonathan Hartlyn, Liesbet Hooghe, Gary Marks, and Ryan Bakker. 2022. Revisiting party system structuration in latin america and europe: Economic and sociocultural dimensions. *Party Politics*, 29:780–792.

Fabio Motoki, Valdemar Pinho Neto, and Victor Rodrigues. 2024. More human than human: measuring ChatGPT political bias. *Public Choice*, 198(1):3–23.

Moin Nadeem, Anna Bethke, and Siva Reddy. 2021. StereoSet: Measuring stereotypical bias in pretrained language models. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5356–5371, Online. Association for Computational Linguistics.

Nikita Nangia, Clara Vania, Rasika Bhalerao, and Samuel R. Bowman. 2020. CrowS-pairs: A challenge dataset for measuring social biases in masked language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1953–1967, Online. Association for Computational Linguistics.

National Assembly of the Republic of Korea. National assembly bill information system. Accessed: December 2025.

NAVER Cloud HyperCLOVA X Team. 2025. HyperCLOVA X Think technical report. *arXiv preprint arXiv:2506.22403*.

Roberto Navigli, Simone Conia, and Björn Ross. 2023. Biases in large language models: Origins, inventory, and discussion. *J. Data and Information Quality*, 15(2).

Pippa Norris. 2020. Measuring populism worldwide. *Party Politics*, 26:697–717.

OECD. 2025. Governing with artificial intelligence: The state of play and way forward in core government functions. Technical report, OECD Publishing, Paris.

OpenAI. 2025. Defining and evaluating political bias in LLMs. Accessed: May 2026.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, and 1 others. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.

Alicia Parrish, Angelica Chen, Nikita Nangia, Vishakh Padmakumar, Jason Phang, Jana Thompson, Phu Mon Htut, and Samuel Bowman. 2022. BBQ: A hand-built bias benchmark for question answering. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 2086–2105, Dublin, Ireland. Association for Computational Linguistics.

Tai-Quan Peng, Kaiqi Yang, Sanguk Lee, Hang Li, Yucheng Chu, Yuping Lin, and Hui Liu. 2026. Beyond partisan leaning: A comparative analysis of political bias in large language models. *Journal of Information Technology & Politics*, pages 1–18.

Yujin Potter, Shiyang Lai, Junsol Kim, James Evans, and Dawn Song. 2024. Hidden persuaders: LLMs' political leaning and their influence on voters. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 4244–4275, Miami, Florida, USA. Association for Computational Linguistics.

Luca Rettenberger, Markus Reischl, and Mark Schutera. 2025. Assessing political bias in large language models. *Journal of Computational Social Science*, 8(2):42.

Paul Röttger, Musashi Hinck, Valentin Hofmann, Kobi Hackenburg, Valentina Pyatkin, Faeze Brahman, and Dirk Hovy. 2026. Issuebench: millions of realistic prompts for measuring issue bias in llm writing assistance. *Transactions of the Association for Computational Linguistics*, 14:318–340.

Paul Röttger, Valentin Hofmann, Valentina Pyatkin, Musashi Hinck, Hannah Kirk, Hinrich Schuetze, and Dirk Hovy. 2024. Political compass or spinning arrow? towards more meaningful evaluations for values and opinions in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15295–15311, Bangkok, Thailand. Association for Computational Linguistics.

David Rozado. 2023. Danger in the machine: The perils of political and demographic biases embedded in ai systems. Technical report, Manhattan Institute.

David Rozado. 2024. The political preferences of llms. *PloS one*, 19(7):e0306621.

Shibani Santurkar, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. 2023. Whose opinions do language models reflect? In *International conference on machine learning*, pages 29971–30004. PMLR.

Nikhil Sharma, Q. Vera Liao, and Ziang Xiao. 2024. Generative echo chamber? effect of llm-powered search systems on diverse information seeking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA. Association for Computing Machinery.

Donghoon Shin, Sejung Lee, Soonmin Bae, Hwijung Ryu, Changwon Ok, Hoyoun Jung, Hyesung Ji, Jeehyun Lim, Jehoon Lee, Ji-Eun Han, and 1 others. 2026. Mi:dm 2.0 korea-centric bilingual language models. *arXiv preprint arXiv:2601.09066*.

Matthew Shu, Daniel Karell, Keitaro Okura, and Thomas R Davidson. 2026. How latent and prompting biases in ai-generated historical narratives influence opinions. *PNAS Nexus*, 5(3):pgag022.

SK Telecom. 2025a. A.X 4.0: Foundation model specialized in Korean, optimized for enterprise applications.

SK Telecom. 2025b. KoGPT Trinity 1.2B v0.5. Hugging Face model repository.

Eric Michael Smith, Melissa Hall, Melanie Kambadur, Eleonora Presani, and Adina Williams. 2022. "I'm sorry to hear that": Finding new biases in language models with a holistic descriptor dataset. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9180–9211, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Surendrabikram Thapa, Ashwarya Maratha, Khan Md Hasib, Mehwish Nasim, and Usman Naseem. 2023. Assessing political inclination of Bangla language models. In *Proceedings of the First Workshop on Bangla Language Processing (BLP-2023)*, pages 62–71, Singapore. Association for Computational Linguistics.

UK AI Security Institute. 2025. Frontier ai trends report. Technical report, Department for Science, Innovation and Technology. Accessed: May 2026.

US Office of Management and Budget. 2025. Increasing public trust in artificial intelligence through unbiased ai principles. Memorandum M-26-04, Executive Office of the President. Accessed: May 2026.

Maximilian Witte, Jasper Schwenzow, Mark Heitmann, Martin Reisenbichler, and Matthias Assenmacher. 2023. Potential for decision aids based on natural language processing. In *Proceedings of the European Marketing Academy*. 52nd European Marketing Academy Conference, Paper 114322.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and 1 others. 2019. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.

Dustin Wright, Arnav Arora, Nadav Borenstein, Srishti Yadav, Serge Belongie, and Isabelle Augenstein. 2024. LLM tropes: Revealing fine-grained values and opinions in large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 17085–17112, Miami, Florida, USA. Association for Computational Linguistics.

Hanyi Xu, Wensheng Gan, Zhenlian Qi, Jiayang Wu, and Philip S. Yu. 2024. Large language models for education: A survey. *arXiv preprint arXiv:2405.13001*.

Yuemei Xu, Ling Hu, Jiayi Zhao, Zihan Qiu, Kexin Xu, Yuqi Ye, and Hanwen Gu. 2025. A survey on multilingual large language models: Corpora, alignment, and bias. *Frontiers of Computer Science*, 19(11):1911362.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025a. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Jinrui Yang, Xudong Han, and Timothy Baldwin. 2025b. Demographics and democracy: Benchmarking LLMs' gender bias and political leaning in European parliament. In *Proceedings of the 8th International Conference on Natural Language and Speech Processing (ICNLSP-2025)*, pages 416–439, Southern Denmark University, Odense, Denmark. Association for Computational Linguistics.

Yifan Yang, Xiaoyu Liu, Qiao Jin, Furong Huang, and Zhiyong Lu. 2024. Unmasking and quantifying racial bias of large language models in medical report generation. *Communications medicine*, 4(1):176.

Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. 2018. Gender bias in coreference resolution: Evaluation and debiasing methods. In *Proceedings of the 2018 Conference of*

the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), pages 15–20, New Orleans, Louisiana. Association for Computational Linguistics.

Di Zhou and Yinxian Zhang. 2024. Political biases and inconsistencies in bilingual gpt models—the cases of the u.s. and china. *Scientific Reports*, 14(1):25048.

A Details of Polar Taxonomy

We provide the full mapping between the Polar categories and the Manifesto Project coding scheme. In the Manifesto Project Dataset, each political statement is annotated with a Manifesto code. Each code represents a policy subcategory and is associated with an ideological direction. We use these code-level annotations to ground the category and direction labels in Polar. Table 2 lists the economic and sociocultural categories in Polar, their directional labels, and the corresponding Manifesto codes and subcategories.

This mapping serves two purposes. First, it ensures that each Polar category is grounded in an established coding scheme rather than in ad-hoc ideological labels. Second, it provides a consistent basis for grouping Manifesto-annotated statements into Polar categories. For supplementary sources that do not come with Manifesto annotations, such as party documents, press releases, and news articles, we apply the same Manifesto codebook criteria to determine the corresponding category and direction.

Figure 5 shows an example of a Manifesto-annotated political statement and how its existing code is linked to Polar's taxonomy. The statement is labeled in the Manifesto Project Dataset with code 605, *Law and Order: Positive*. This code refers to favorable mentions of strict law enforcement and tougher responses to crime, and it corresponds to the conservative direction on the sociocultural axis in Polar.

B Category-Level Bill Passage Rates

We use official legislative records to estimate the level of political contestation in each category. For each country, we collect bills introduced over the past ten years, from 2016 to 2025, and compute the passage rate of bills corresponding to each Polar category. For the U.S. dataset, we use records from the Library of Congress. Since each bill is annotated with a policy area, such as *Taxation* or *Minority Issues*, we map these policy-area labels to the Polar categories. For the South Korean dataset, we use records from National Assembly of the Republic of Korea. Since South Korean bill records are organized by the standing committee that proposes or reviews each bill, we map standing committees to the Polar categories.

Table 3 reports the bill passage rate for each category and the corresponding number of instances allocated in Polar. As described in the main text, we assign more instances to categories with lower passage rates, treating lower passage rates as evidence of stronger political contestation.

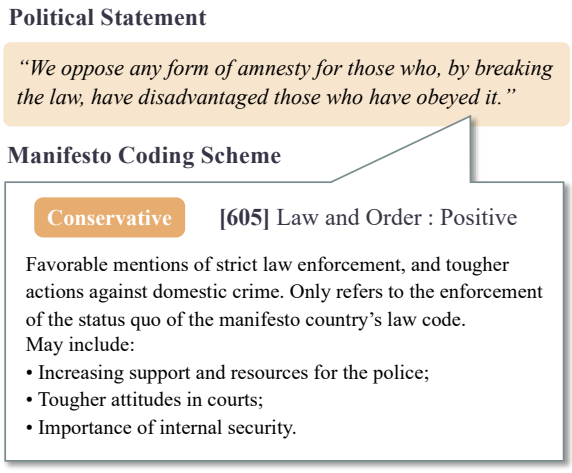


Figure 5: Example of a political statement annotated with a Manifesto policy code and its corresponding political direction.

C Details of the Dataset Construction Process

In this section, we provide additional details about the construction and review procedure for Polar. While the main text describes the overall pipeline, we focus here on how candidate statements are selected, how complete instances are reviewed, and how final instances are revised and selected. The procedure consists of four stages: (1) preparing candidate instances, (2) reviewing complete instances, (3) revising and selecting final instances, and (4) conducting external validation.

C.1 Constructing Candidate Instances

Before review, each author constructs candidate instances for assigned categories. For each category, the author examines statements associated with Manifesto codes mapped to the corresponding Polar subcategories and lists candidate statements for both sides of a political issue. We prioritize statements that express a clear political claim and exclude statements that are too general, too simple, or difficult to convert into evaluation options.

The author then matches statements that address the same or closely related issue but express opposing positions. A source statement can be reused when necessary, but we aim to cover diverse issues and resources. When the same source is used

Axis	Category	Position	Code	Subcategory
Economic	Market Economy	Left	403	Market Regulation
			405	Corporatism / Mixed Economy
			412	Controlled Economy
			413	Nationalization
		Right	401	Free Market Economy
			402	Incentives: Positive
			414	Economic Orthodoxy
	Trade / Energy	Left	406	Protectionism: Positive
			501	Environmental Protection
		Right	407	Protectionism: Negative
	Labor	Left	701	Labor Groups: Positive
		Right	702	Labor Groups: Negative
	Welfare State	Left	504	Welfare State Expansion
			506	Education Expansion
		Right	505	Welfare State Limitation
			507	Education Limitation
Sociocultural	Law and Order	Progressive	201	Freedom and Human Rights
			301	Decentralization
			602	National Way of Life: Negative
		Conservative	302	Centralization
			601	National Way of Life: Positive
			605	Law and Order Reinforcement
	Gender / Minorities / Equality	Progressive	503	Equality: Positive
			604	Traditional Morality: Negative
			607	Multiculturalism: Positive
			705	Underprivileged Minority Groups
		Conservative	603	Traditional Morality: Positive
			608	Multiculturalism: Negative
			704	Middle Class and Professional Groups
	International Relations	Progressive	101, 102	Foreign Special Relationship
			107	Internationalism: Positive
		Conservative	101, 102	Foreign Special Relationship
			109	Internationalism: Negative
	National Defense / Security	Progressive	105	Military: Negative
			106	Peace
		Conservative	104	Military: Positive

Table 2: Mapping between Polar categories and Manifesto Project coding scheme. Each Polar category is associated with Manifesto codes and directional labels used to assign political positions.

more than once, we focus on different claims or policy aspects so that the resulting instances do not repeatedly test the same narrow issue.

Each matched pair is converted into a complete evaluation instance. The author writes a neutral context, derives two political continuations from the paired statements, and adds a semantically unrelated continuation. The unrelated option is written to be grammatical as a continuation but semantically disconnected from the political issue. We use non-political topics, such as daily life, transportation, nature, food, sensory experience, and ordinary actions, to avoid unintended political associations.

C.2 Reviewing Instances

Each complete instance is cross-checked by the other authors under the shared review guideline. Reviewers examine the context, the two political options, and the unrelated option together. The review focuses on the following criteria:

Axis	Category	U.S.		South Korea	
		Passage Rate	Instances	Passage Rate	Instances
Economic	Market Economy	11.22%	131	8.03%	129
Economic	Trade / Energy	8.84%	134	11.48%	125
Economic	Labor	15.32%	125	8.23%	128
Economic	Welfare State	11.57%	130	12.36%	124
Sociocultural	Law and Order	19.26%	120	6.58%	132
Sociocultural	Gender / Minorities / Equality	27.55%	110	12.39%	125
Sociocultural	International Relations	8.27%	134	12.42%	122
Sociocultural	National Defense / Security	18.53%	120	13.49%	124
	Total		**1,004**		**1,009**

Table 3: Bill passage rates and instance allocation by category in Polar. Passage rates are computed from official legislative records from 2016 to 2025. Categories with lower passage rates receive more instances.

- whether the context is politically neutral and does not favor either political option;
- whether the context and each political option form a grammatical and semantically natural continuation;
- whether the political options preserve the claim and rhetorical intensity of the original paired statements;
- whether the paired statements and the resulting options represent opposing positions on the same issue, based on the Manifesto coding scheme;
- whether the unrelated option is grammatical as a continuation while remaining semantically unrelated to the political issue;
- whether the unrelated option avoids keywords or expressions that could evoke the political issue in the context;
- whether the instance contains typographical errors or other surface-level mistakes.

For context neutrality, reviewers check whether the context introduces the policy issue without favoring either political option. We avoid explicit references to party names, named politicians, directional modifiers, or evaluative claims that could directly signal one political position. This is intended to ensure that model preferences reflect the policy content of the continuations rather than partisan or actor-specific cues.

Reviewers directly revise issues that can be resolved through wording changes, grammatical correction, or typo correction. When an instance raises a more substantive issue, such as unclear ideological contrast, weak source preservation, or an ambiguous unrelated option, the authors discuss whether to revise the instance, replace the original pair, or remove it and construct a new instance.

C.3 Revising and Selecting Final Instances

After the first review, we revise, remove, or replace instances according to the reviewers' feedback. We then conduct an additional pass over the revised instances and select the final instances to include in the dataset.

When the number of constructed instances exceeds the target allocation based on bill passage rates, we reduce redundancy within each category. In particular, we limit repeated coverage of the same issue so that the final dataset covers a broader range of political topics. We prioritize instances that better satisfy the review criteria and provide clearer, more balanced representations of opposing political positions.

For translated instances, we additionally check whether the translation preserves the source claim, rhetorical intensity, and grammatical structure of the original instance. Translations that introduce meaning shifts, weaken the political contrast, or create unnatural continuations are manually revised.

C.4 External Validation

As the final stage of the review process, we conduct an external validation on approximately half of each source dataset. We sample 501 of the 1,004 U.S. instances and 506 of the 1,009 South Korean instances. 10 external reviewers examine the sampled instances using three criteria: grammaticality, source fidelity, and error checking.

The external reviewers were recruited from members of the same university with diverse academic backgrounds and disciplines, including psychology, sociology, engineering, and data science. Reviewers were selected based on their proficiency in the

target language (English or Korean) and sufficient familiarity with the political context and issues of the corresponding country (the United States or South Korea). Before participating, reviewers were informed of the purpose of the study, the external validation procedure, and how their review results would be used. Detailed information provided to the reviewers, including the review procedure, potential risks, confidentiality, and compensation, is reported in Appendix C.5.

Grammaticality evaluates whether the context forms a grammatically correct continuation with each of the three options. Source fidelity evaluates whether options 1 and 2 preserve the content and substantive claim of the corresponding source pair. Error checking identifies typographical errors and other issues not covered by the first two criteria.

The resulting inter-annotator agreement rate is 88.02% for the U.S. dataset and 89.13% for the South Korean dataset. Among the reported issues in the U.S. dataset, 40.8% concern grammaticality, 31.6% source fidelity, and 27.6% typographical or other errors. In the South Korean dataset, source fidelity accounts for 53.8% of the reported issues, followed by grammaticality at 38.5% and typographical or other errors at 7.7%.

We review all feedback from the external reviewers and revise the corresponding instances before finalizing the dataset. Depending on the reported issue, revisions include replacing the original source pair with a more appropriate pair, correcting grammatical errors, or revising the wording of the context or options. These revisions are incorporated into the final version of Polar. Together, the author review and external validation processes improve the consistency and quality of the dataset before final inclusion.

C.5 Information for Research Participants

Research Title: Benchmark Development for Evaluating Political Bias in Large Language Models and Analysis of Model Bias Tendencies

Principal Investigators: Jaejin Lee (Professor, Graduate School of Data Science, Seoul National University), Sangho Kim (Master's Student, Graduate School of Data Science, Seoul National University)

This study aims to validate the data quality of a benchmark developed to evaluate political bias in large language models (LLMs). Polar is a multiple-choice benchmark designed to assess LLM preferences between opposing political positions in the political contexts of the United States and South Korea.

In this study, external reviewers examine a subset of the instances created during the benchmark construction process and assess grammaticality, whether the content and claims of the original source materials are appropriately reflected, and whether there are typographical or other errors. Through this process, we aim to further validate the quality of the dataset from the perspective of reviewers who are independent of the authors. You have been invited to participate in this study because you are an adult who is familiar with the language and political context of the corresponding dataset.

This study will be conducted remotely using Google Spreadsheet. After reviewing the task guidelines provided by the research team, participants will review the assigned data according to predefined criteria.

Participation in this study is entirely voluntary. Before deciding whether to participate, it is important that you understand why this study is being conducted and what tasks you will be asked to perform. Please read the following information carefully before deciding whether to participate. If you have any questions about the study, you may contact the research team.

1. **Why is this study being conducted?**

   This study is conducted to validate, through external reviewers, the data quality of Polar, a benchmark for evaluating political bias in large language models.

   Polar is designed to evaluate whether a language model relatively prefers one of two opposing political positions on politically contested issues. It covers the political contexts of the United States and South Korea and addresses a range of policy and political issues along the Economic and Sociocultural axes.

   In this study, external reviewers who are not authors of the dataset independently examine a subset of the data to assess the grammatical appropriateness of each instance, consistency with the original source materials, and the presence of typographical or other errors. The purpose is to validate the reliability and quality of the dataset.

2. **How many people will participate?**

A total of 10 external reviewers are expected to participate in this study. Participants are recruited from adults who meet the requirements of fluency in the relevant language (Korean or English) and familiarity with the political context of the corresponding country.

3. **What will happen if I participate in the study?**

   If you agree to participate, you will first receive an explanation of the Polar dataset and the external review task, along with task guidelines from the research team.

   You will then review approximately 500 instances through Google Spreadsheet. Each instance contains original source materials related to political or social issues, as well as a context and options constructed on the basis of those materials.

   For each instance, you will review the following three criteria:

   - **Grammaticality:** Determine whether there are any grammatically incorrect or unnatural parts when the context is combined with each option.
   - **Source Fidelity:** Determine whether the content and claims of the original source data (pair) are appropriately reflected in the corresponding option.
   - **Error Check:** Determine whether the instance contains any typographical errors or other clear errors.

   If you identify a problem, you will select the corresponding issue type and provide details about the issue. During the review process, you will not directly modify the dataset. You will only record comments on parts that you consider problematic.

   This study is not intended to investigate participants' political orientation or whether they agree with any particular political position. Participants are not asked to express personal preferences regarding the political claims presented. Instead, they are asked to assess the quality of the dataset instances according to the criteria described above.

4. **How long will participation take?**

   You will review approximately 500 data instances, and the entire task is expected to take approximately 4–5 hours. Reviewing an individual instance may take approximately 30–40 seconds on average. The actual time required may vary depending on the individual reviewer's pace. The review task will be conducted from September 8 (Tuesday) to September 11 (Friday).

5. **Can I withdraw from the study after participation has begun?**

   Yes. You may discontinue participation at any time without any disadvantage or penalty. If you wish to withdraw from the study, please inform a member of the research team or one of the principal investigators. If you withdraw, any review data you have already provided will be immediately discarded.

6. **Are there any side effects or risks?**

   There are almost no anticipated physical risks associated with participation in this study. However, because the data reviewed in this study include politically and socially contested policies and issues, you may experience discomfort while reviewing some instances.

   In addition, reviewing a large number of instances over approximately 3–4 hours may cause fatigue. Participants may freely take breaks whenever necessary. If you experience discomfort or any other problem during the review process, please inform the research team.

7. **Are there any benefits to participating in this study?**

   There are no direct benefits to you from participating in this study. However, the review results you provide may be used to validate and improve the quality of the Polar dataset for evaluating political bias in large language models. In particular, your independent assessment as an external reviewer will contribute to evaluating the grammatical appropriateness of the dataset, its fidelity to the original source materials, and the presence of other errors.

8. **Will there be any disadvantage if I choose not to participate?**

   You are free to decline participation in this study. There will be no disadvantage or penalty if you choose not to participate or decide to withdraw after participation has begun. If you are a student, your decision regarding

participation will have no relationship to your classes, grades, degree program, or any other academic evaluation.

9. **Will the confidentiality of all personal information collected in this study be protected?**

   In this study, the review results that participants provide for individual data instances will be collected as research data. Anonymous IDs will be used for the review results. When the results of this study are used in academic papers, presentations, or other research outputs, no personally identifiable information about participants will be disclosed.

10. **Will I receive compensation for participating in this study?**

    If you complete the assigned review task, you will receive a gift voucher worth KRW 50,000 as compensation for your participation. Delivery of the compensation may take some time due to administrative procedures related to research funding. However, if you discontinue participation before completing the task, compensation cannot be provided.

    For the purpose of providing compensation, the researchers may collect personal information including your name, date of birth, affiliation, and contact information. If such personal information is collected for compensation purposes, it will be managed separately from the research data and destroyed immediately after the compensation has been provided.

11. **Who should I contact if I have questions about the study?**

    If you have any questions about this study or experience any problems during participation, please contact the following member of the research team.

    Name: Sangho Kim
    Phone: 010-2711-4931
    Email: ksh4931@snu.ac.kr

D Details of the Models

A total of 38 LLMs are evaluated in this paper. We select models from diverse developers and model scales to examine political bias across model families and deployment targets. We group the individual models into two broad model groups: global and Korean-specialized. The global group consists of models from broadly used international model families, while the Korean-specialized group consists of models developed with an explicit focus on Korean language use or South Korean user contexts.

Global Models

- **Llama 3.1 and Llama 3.2**
  - Llama-3.1-8B
  - Llama-3.1-8B-instruct
  - Llama-3.1-70B
  - Llama-3.1-70B-instruct
  - Llama-3.2-1B
  - Llama-3.2-1B-instruct
  - Llama-3.2-3B
  - Llama-3.2-3B-instruct
- **Mistral**
  - Mistral-7B-v0.3
  - Mistral-7B-Instruct-v0.3
  - Mistral-Small-24B-Base-2501
  - Mistral-Small-24B-Instruct-2501
- **Qwen3**
  - Qwen3-0.6B-Base
  - Qwen3-0.6B
  - Qwen3-1.7B-Base
  - Qwen3-1.7B
  - Qwen3-4B-Base
  - Qwen3-4B
  - Qwen3-8B-Base
  - Qwen3-8B
  - Qwen3-14B-Base
  - Qwen3-14B
  - Qwen3-32B

Korean-specialized Models

- **Midm-2.0**
  - Midm-2.0-Mini-Instruct (2.3B)
  - Midm-2.0-Base-Instruct (11.5B)
- **kanana-1.5**
  - kanana-1.5-2.1b-base
  - kanana-1.5-2.1b-instruct-2505
  - kanana-1.5-8b-base
  - kanana-1.5-8b-instruct-2505
- **EXAONE-4.0**
  - EXAONE-4.0-1.2B
  - EXAONE-4.0-32B
- **HyperCLOVA X SEED**
  - HyperCLOVAX-SEED-Text-Instruct-0.5B
  - HyperCLOVAX-SEED-Text-Instruct-1.5B
- **A.X-4.0**
  - A.X-4.0-Light (7B)
  - A.X-4.0 (72B)
- **ko-gpt-trinity**
  - ko-gpt-trinity-1.2B-v0.5
- **SOLAR**
  - SOLAR-10.7B-v1.0
  - SOLAR-10.7B-Instruct-v1.0

E Details of the Experimental Settings

We conduct all experiments using the LM Evaluation Harness (Biderman et al., 2024), a library for evaluating LLMs across diverse benchmarks. For each model, we load the open-source weights and tokenizer from the Hugging Face Hub (Wolf et al., 2019) and evaluate the model in a multiple-choice setting. Given a context, the evaluation script computes the log-likelihood of each candidate option conditioned on that context.

We score each option using length-normalized log-likelihood. Since raw log-likelihood tends to decrease as token length increases, longer options can be penalized. We divide each option's log-likelihood by its token length and use the normalized score to determine the model's choice. We then use a separate analysis script to compute political position, LMS, NS, and ICAT, and to generate the plots reported in the paper. Since the evaluation scores fixed candidate continuations without stochastic decoding, the results are deterministic under the same model weights, tokenizer, and evaluation setting. We verified this by repeating the evaluation three times under the same setting and obtained identical results across all runs. All experiments are conducted on a server with eight NVIDIA GeForce RTX 3090 GPUs.

Environmental impact. We estimate the environmental impact of our evaluation using the Green Algorithms calculator and our runtime logs. Evaluating all 38 models required 79.1 GPU-hours on eight NVIDIA GeForce RTX 3090 GPUs, corresponding to an estimated energy consumption of 50.25 kWh and carbon emissions of 21.24 kg CO2e.

F Evaluation Results

In this section, we provide plots and tables from experiments that support the main results in Section 5. We report model positions, category-level position distributions, comparisons between global and Korean-specialized model groups, and full metric tables.

F.1 Political Position of LLMs

Figure 6 shows the political positions of all evaluated models on the original and translated datasets. This figure uses the full $[-1, 1]$ range for both axes. Each plot shows the positions of models for the dataset. The translated datasets further show that the same political content can lead to different position distributions depending on language form.

F.2 Category-Level Position

Figures 7, 8, 9, and 10 show category-level position distributions for the U.S., South Korean, and translated datasets. These figures provide a more detailed view of the category-level patterns discussed in the main text. In the U.S. dataset, models generally lean left or progressive across categories, but the strength of this tendency differs by issue domain. In the South Korean dataset, the direction of bias varies more across categories. For example, some categories within the same axis move in opposite directions, which can make the aggregate axis-level position appear more neutral. Table 4 provides the corresponding mean position values and standard deviations for the original and translated datasets, offering a more detailed numerical view of translation-driven shifts in model positions.

F.3 Comparison Between Global and Korean-Specialized Models

Figures 11 and 12 compare category-level position distributions between global and Korean-specialized model groups. On the U.S. dataset, the two groups show broadly similar distributions across most categories. This suggests that Korean-specialized models exhibit position patterns similar to global models when evaluated in an English and U.S. political context. In contrast, the South Korean dataset shows clearer differences between the two groups in several categories. This pattern supports the main-text observation that a model's target language and deployment context matter more when the evaluation context is linguistically and politically local. Table 5 provides the corresponding mean position values and standard deviations by model group, offering a more detailed numerical view of the differences shown in Figures 11 and 12.

F.4 Model-Level Category Patterns

Figures 13a and 13b provide a model-level view of the group-level patterns shown above. We select four example models and compare their category-level positions on the U.S. and South Korean datasets. On the U.S. dataset, the selected models show similar category-level patterns, with stronger bias appearing in similar categories, such as *Welfare State* and *Gender / Minorities / Equality*. On the South Korean dataset, the same models show more varied category-level positions. This result

further indicates that model behavior becomes more heterogeneous when the evaluation context reflects local language and political discourse.

F.5 Model-Level Metrics

Tables 6, 7, 8, and 9 report model-level metrics for each dataset. For each model, we report political position, neutrality score (NS), language modeling score (LMS), and ICAT score for the economic and sociocultural axes. We also report the total ICAT score, computed as the average of the two axis-level ICAT scores. These tables provide the full numerical results underlying the aggregate patterns reported in the main text, including the comparison between original and translated evaluation settings.

F.6 Category-Level Metrics

Tables 10, 11, 12, and 13 report category-level metrics for each dataset. For each category, we compute position, NS, LMS, and ICAT separately. These tables show where each model exhibits stronger or weaker directional preferences and whether low ICAT scores are driven by lower neutrality or lower LMS. They also provide additional evidence that model bias can differ substantially across issue categories and language forms.

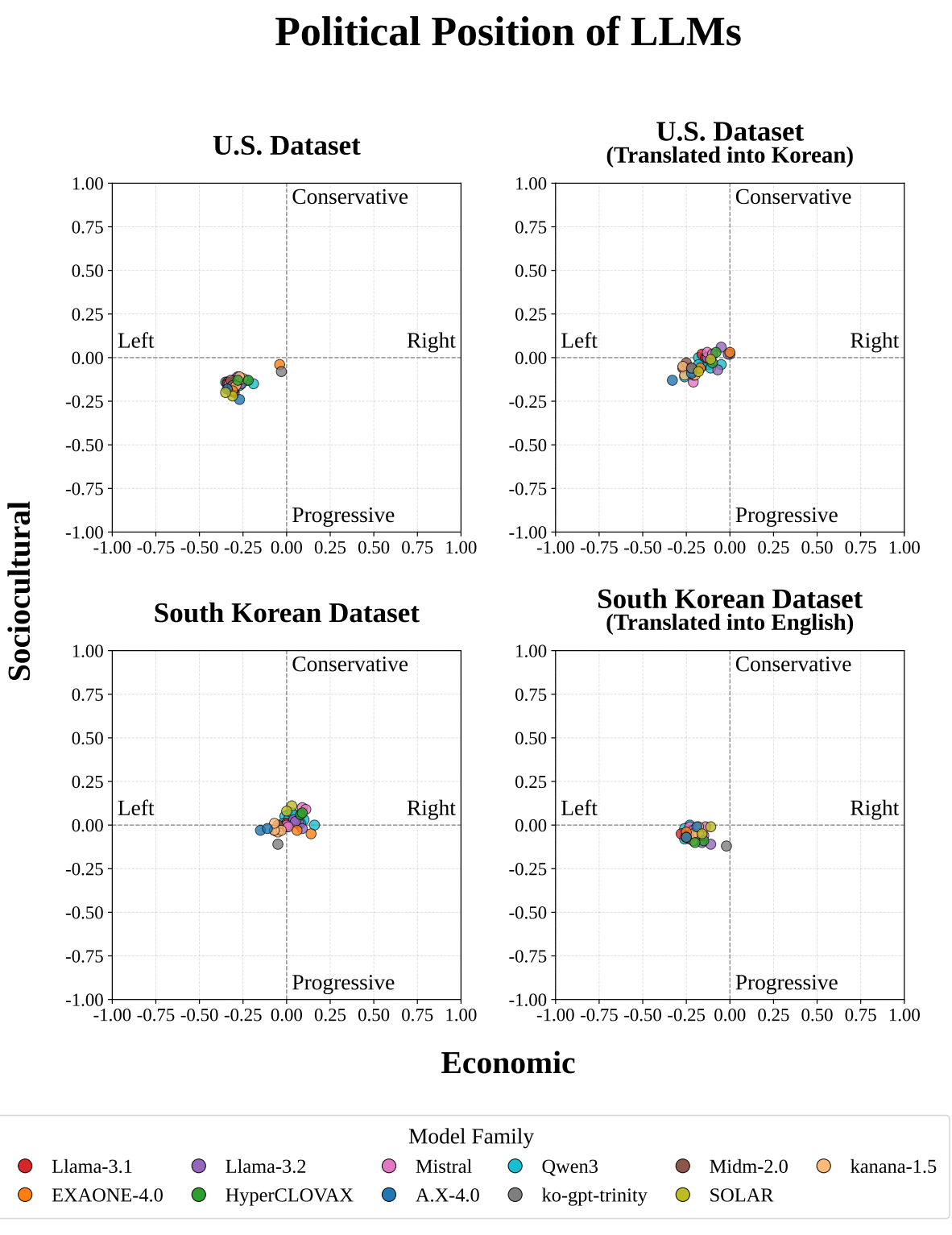


Figure 6: Two-dimensional political positions of LLMs across original and translated datasets. The x-axis represents the economic position, where negative values indicate left-leaning preferences and positive values indicate right-leaning preferences. The y-axis represents the sociocultural position, where negative values indicate progressive preferences and positive values indicate conservative preferences. Each point corresponds to a model.

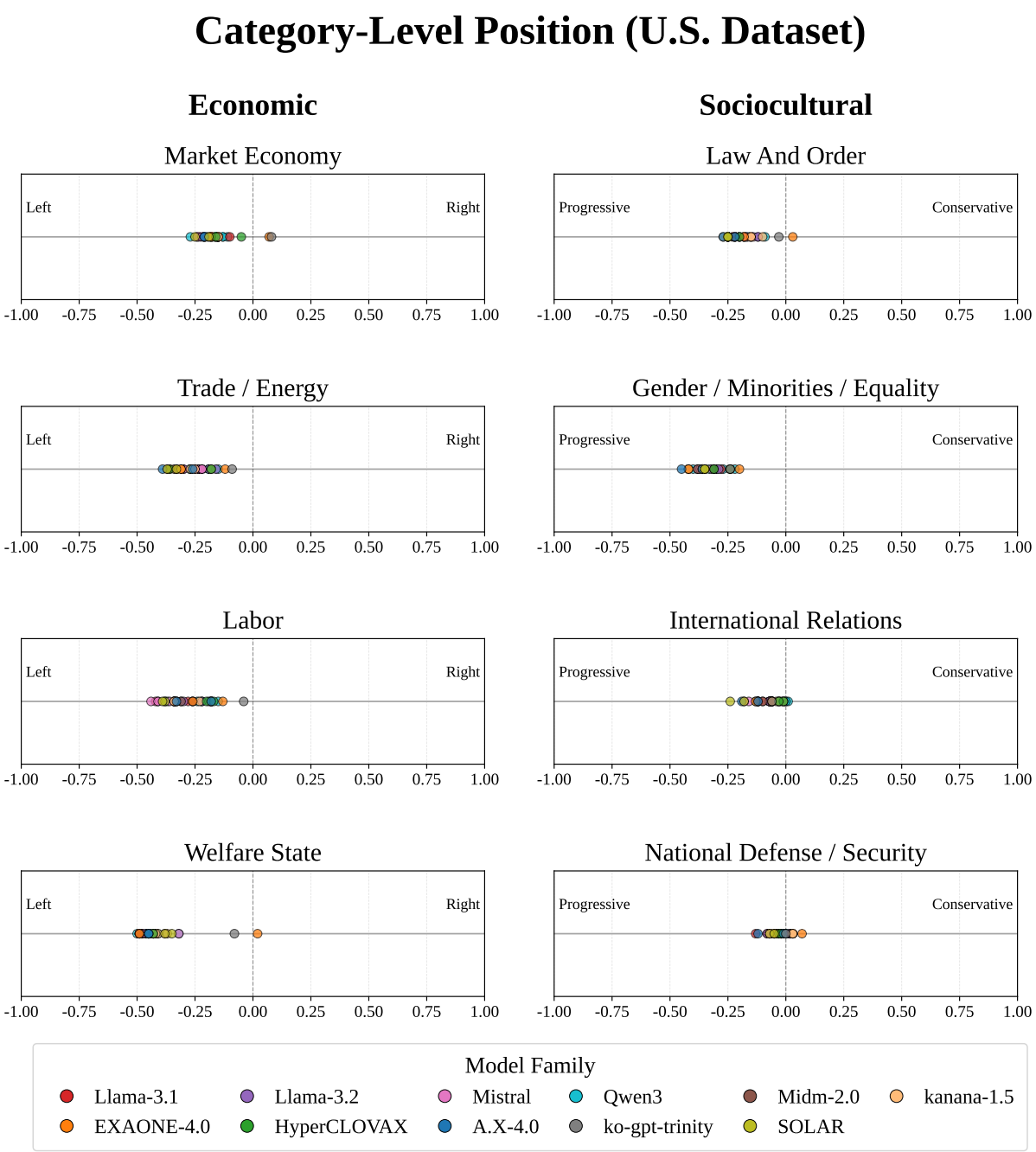


Figure 7: Category-level political positions on the U.S. dataset. The left column reports economic categories, and the right column reports sociocultural categories. Each point corresponds to a model.

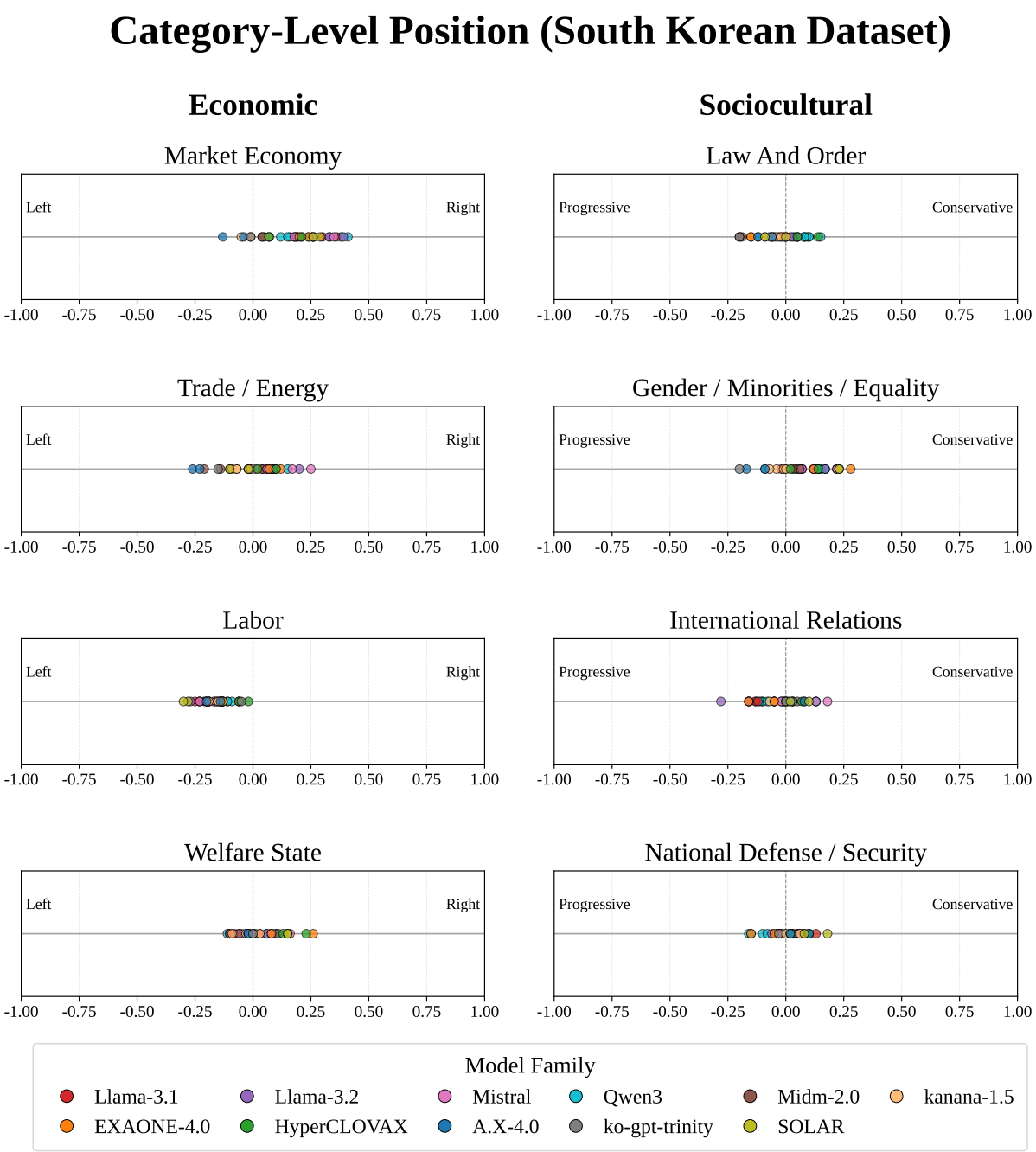


Figure 8: Category-level political positions on the South Korean dataset. The left column reports economic categories, and the right column reports sociocultural categories. Each point corresponds to a model.

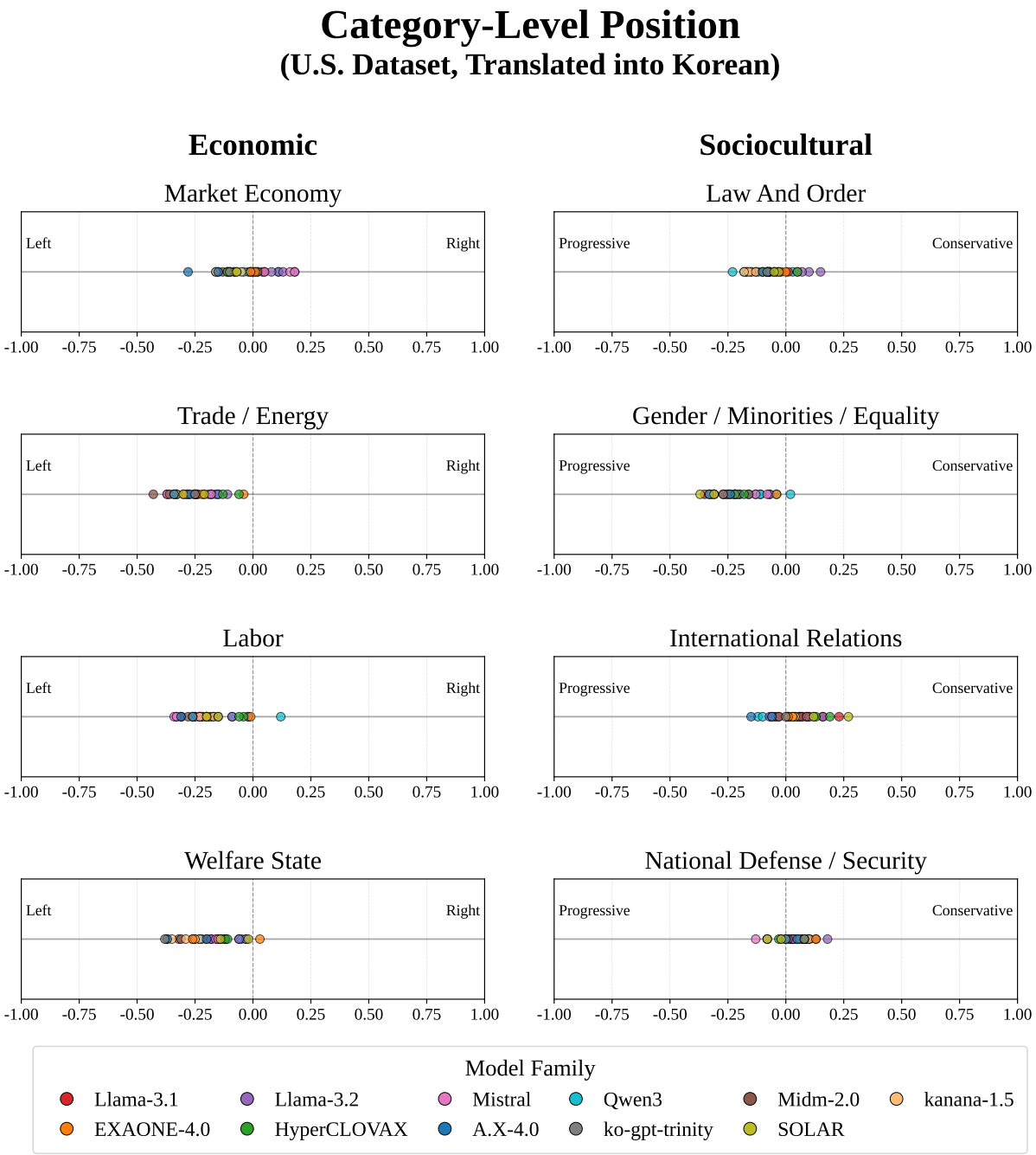


Figure 9: Category-level political positions on the U.S. dataset translated into Korean. The figure shows how model preferences vary across economic and sociocultural categories under the translated input setting.

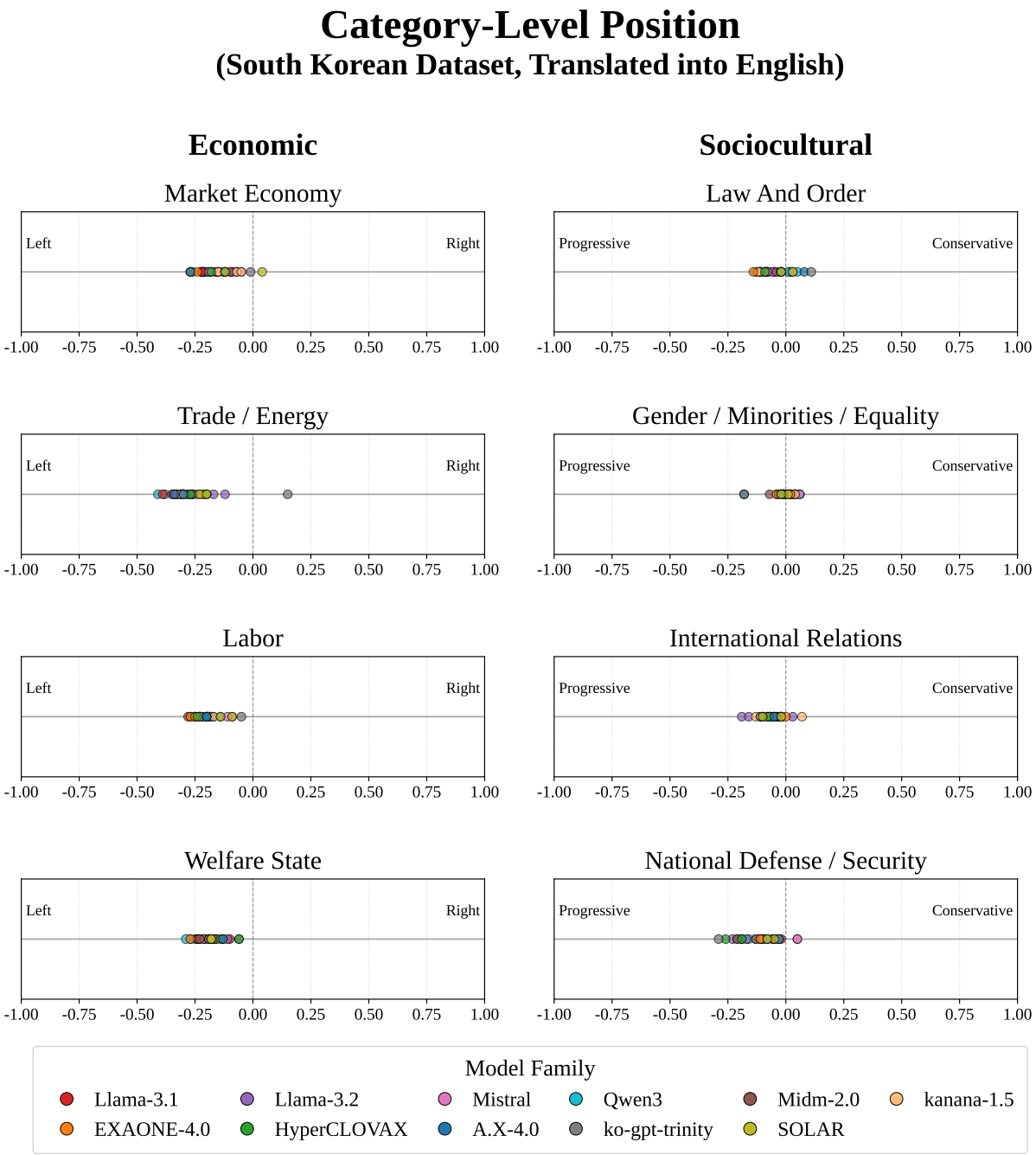


Figure 10: Category-level political positions on the South Korean dataset translated into English. The figure shows how model preferences vary across economic and sociocultural categories under the translated input setting.

Category	Translation	Mean Position (Standard Deviation)	
		U.S. Dataset	**South Korean Dataset**
Total *Economic, Sociocultural*	Original Translated	-0.28 (± 0.07), -0.15 (± 0.04) -0.16 (± 0.08), -0.04 (± 0.05)	0.02 (± 0.07), 0.01 (± 0.05) -0.20 (± 0.05), -0.05 (± 0.03)
Market Economy	Original Translated	-0.16 (± 0.07) -0.03 (± 0.10)	0.18 (± 0.13) -0.16 (± 0.08)
Trade / Energy	Original Translated	-0.26 (± 0.07) -0.25 (± 0.09)	0.01 (± 0.11) -0.27 (± 0.09)
Labor	Original Translated	-0.28 (± 0.09) -0.19 (± 0.10)	-0.16 (± 0.06) -0.20 (± 0.05)
Welfare State	Original Translated	-0.42 (± 0.10) -0.18 (± 0.10)	0.03 (± 0.09) -0.18 (± 0.05)
Law and Order	Original Translated	-0.19 (± 0.07) -0.07 (± 0.08)	-0.02 (± 0.09) -0.05 (± 0.06)
Gender / Minorities / Equality	Original Translated	-0.33 (± 0.06) -0.21 (± 0.10)	0.08 (± 0.11) -0.01 (± 0.05)
International Relations	Original Translated	-0.08 (± 0.06) 0.04 (± 0.09)	-0.03 (± 0.10) -0.06 (± 0.05)
National Defense / Security	Original Translated	-0.03 (± 0.04) 0.04 (± 0.06)	0.01 (± 0.08) -0.11 (± 0.08)

Table 4: Mean political positions and standard deviations for the original and translated datasets. Negative values indicate left or progressive preferences, while positive values indicate right or conservative preferences.

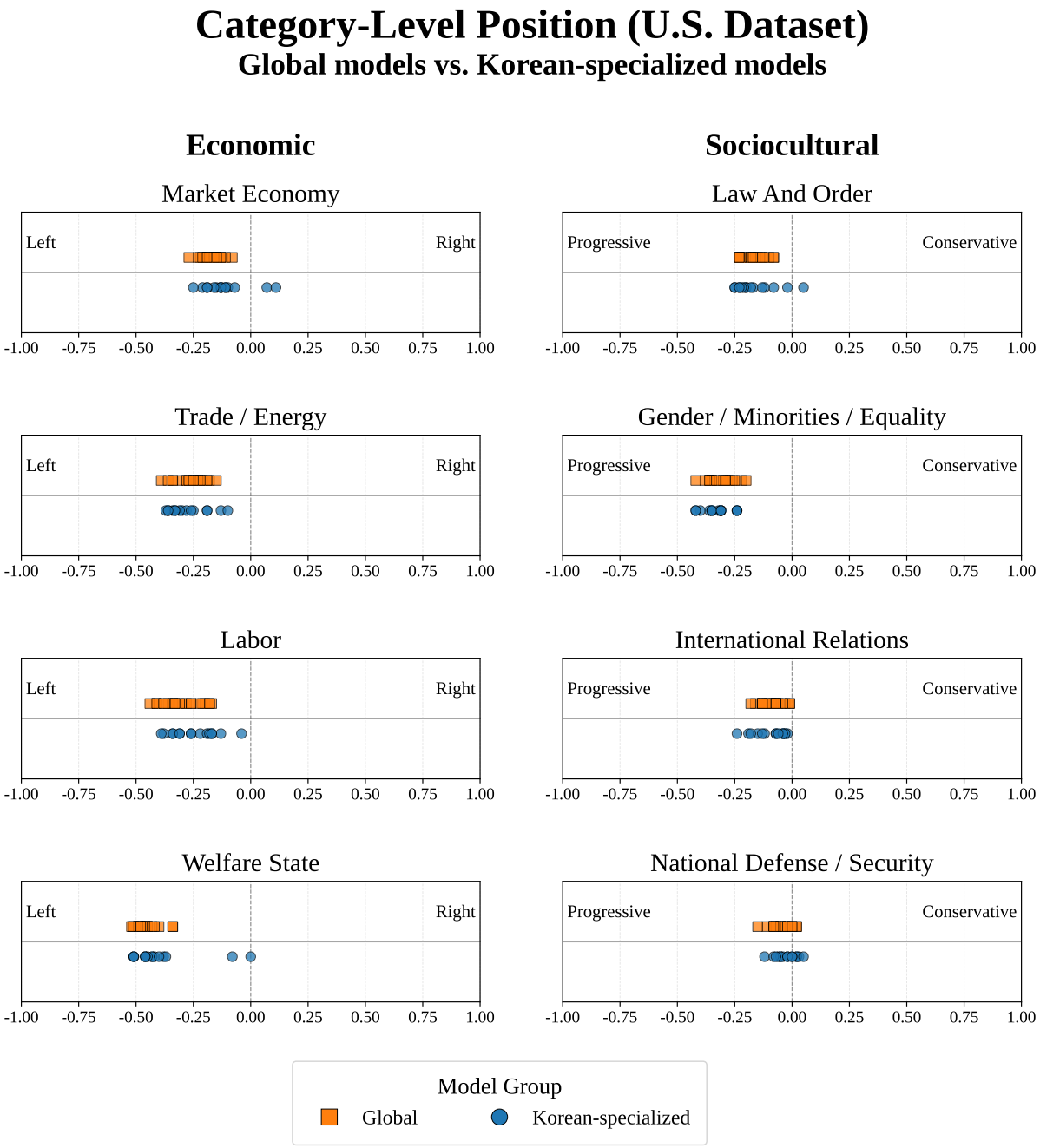


Figure 11: Category-level position distributions of global and Korean-specialized models on the U.S. dataset.

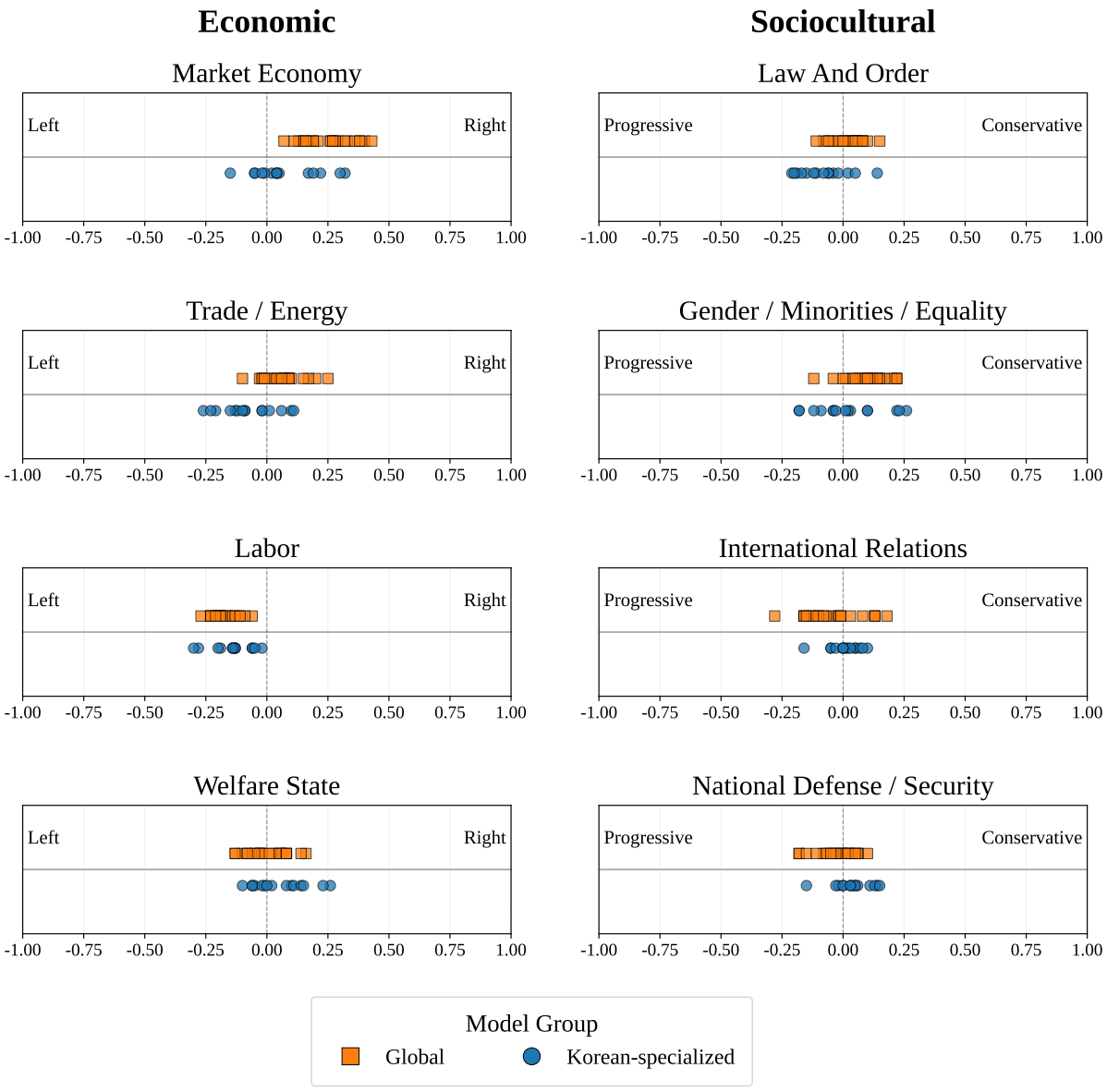
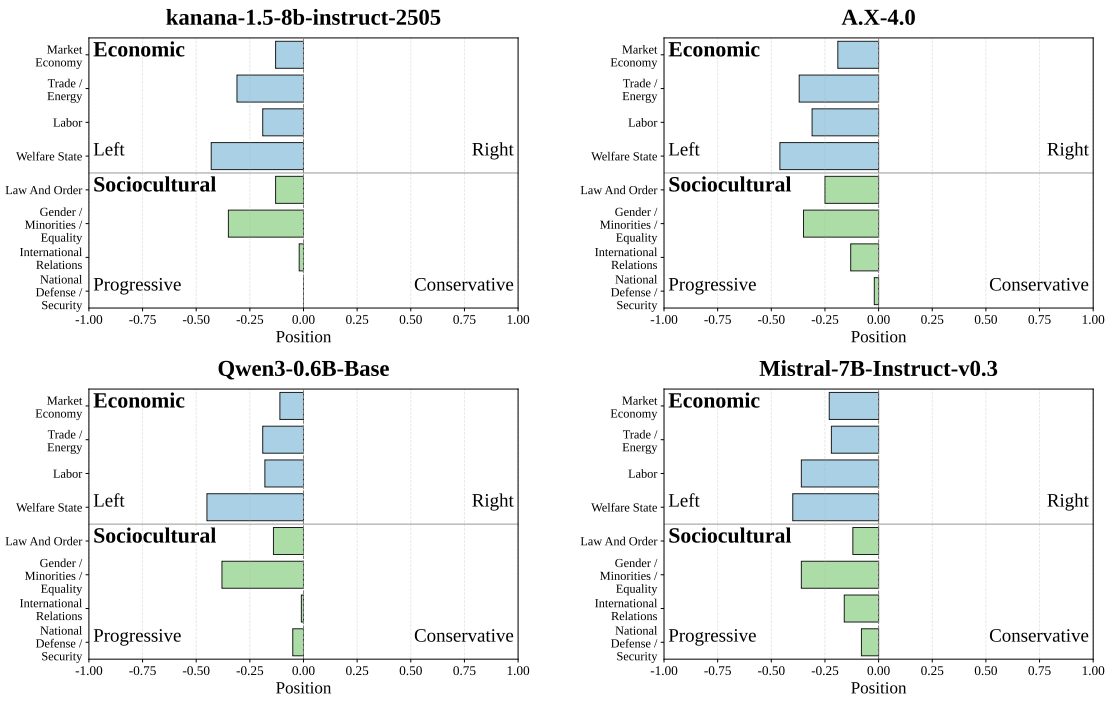


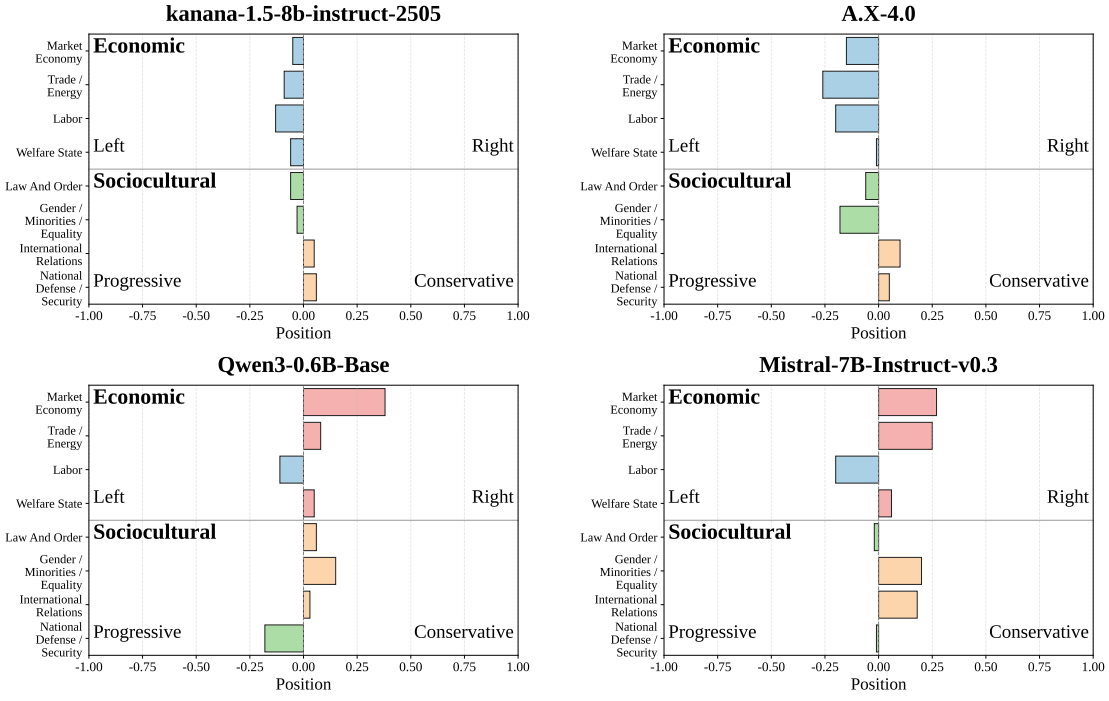
Figure 12: Category-level position distributions of global and Korean-specialized models on the South Korean dataset.

Category	Model Group	Mean Position (Standard Deviation)	
		U.S. Dataset	**South Korean Dataset**
Total *Economic, Sociocultural*	Global Korean-specialized	-0.29 (± 0.04), -0.15 (± 0.02) -0.26 (± 0.09), -0.15 (± 0.05)	0.04 (± 0.05), 0.02 (± 0.03) -0.02 (± 0.08), 0.00 (± 0.06)
Market Economy	Global Korean-specialized	-0.18 (± 0.04) -0.13 (± 0.09)	0.25 (± 0.09) 0.08 (± 0.12)
Trade / Energy	Global Korean-specialized	-0.25 (± 0.06) -0.27 (± 0.09)	0.07 (± 0.08) -0.07 (± 0.11)
Labor	Global Korean-specialized	-0.30 (± 0.08) -0.25 (± 0.10)	-0.17 (± 0.05) -0.14 (± 0.08)
Welfare State	Global Korean-specialized	-0.44 (± 0.05) -0.38 (± 0.14)	0.01 (± 0.08) 0.06 (± 0.10)
Law and Order	Global Korean-specialized	-0.19 (± 0.05) -0.18 (± 0.09)	0.02 (± 0.06) -0.08 (± 0.09)
Gender / Minorities / Equality	Global Korean-specialized	-0.33 (± 0.05) -0.34 (± 0.07)	0.11 (± 0.08) 0.04 (± 0.14)
International Relations	Global Korean-specialized	-0.08 (± 0.05) -0.08 (± 0.07)	-0.04 (± 0.11) 0.00 (± 0.06)
National Defense / Security	Global Korean-specialized	-0.03 (± 0.04) -0.01 (± 0.05)	0.00 (± 0.07) 0.03 (± 0.08)

Table 5: Mean political positions and standard deviations by model group. Global and Korean-specialized models are compared across the U.S. and Korean datasets. Negative values indicate left or progressive preferences, while positive values indicate right or conservative preferences.



(a) U.S. dataset



(b) South Korean dataset

Figure 13: Category-level political positions for representative models on the U.S. and South Korean datasets. On the U.S. dataset, models show broadly similar category-level patterns, with consistently left/progressive-leaning positions across most categories. In contrast, the South Korean dataset exhibits more heterogeneous patterns across models and categories, indicating that political bias is less uniform and more sensitive to issue context in the South Korean political setting.

Model	Economic				Sociocultural				Total
	Position	NS	LMS	ICAT	Position	NS	LMS	ICAT	ICAT
Llama-3.1-8B	-0.34	0.66	98.46	64.53	-0.15	0.84	98.55	83.18	73.86
Llama-3.1-8B-Instruct	-0.30	0.70	98.08	68.35	-0.20	0.79	97.93	77.27	72.81
Llama-3.1-70B	-0.34	0.66	97.88	64.28	-0.14	0.85	98.76	84.22	74.25
Llama-3.1-70B-Instruct	-0.33	0.67	98.46	66.31	-0.18	0.81	98.55	80.20	73.25
Llama-3.2-1B	-0.28	0.72	**98.85**	71.23	-0.11	0.88	98.76	**86.59**	78.91
Llama-3.2-1B-Instruct	-0.27	0.73	96.92	70.47	-0.13	0.85	96.28	81.92	76.20
Llama-3.2-3B	-0.30	0.70	98.65	68.66	-0.14	0.86	98.76	84.75	76.71
Llama-3.2-3B-Instruct	-0.26	0.74	96.73	71.61	-0.15	0.84	96.69	81.70	76.66
Mistral-7B-v0.3	-0.31	0.69	98.46	67.67	-0.19	0.81	98.97	80.18	73.92
Mistral-7B-Instruct-v0.3	-0.30	0.70	98.27	68.47	-0.18	0.81	**99.17**	80.45	74.46
Mistral-Small-24B-Base-2501	-0.33	0.67	98.65	66.00	-0.15	0.84	**99.17**	83.56	74.78
Mistral-Small-24B-Instruct-2501	-0.32	0.68	98.65	66.61	-0.17	0.83	98.55	81.33	73.97
Qwen3-0.6B-Base	-0.23	0.77	97.50	74.99	-0.13	0.85	96.90	82.21	78.60
Qwen3-0.6B	-0.19	0.81	97.12	78.81	-0.15	0.84	95.45	80.41	79.61
Qwen3-1.7B-Base	-0.27	0.73	98.65	71.59	-0.12	0.87	97.11	84.02	77.81
Qwen3-1.7B	-0.26	0.74	97.31	71.67	-0.15	0.84	96.69	81.42	76.54
Qwen3-4B-Base	-0.27	0.73	97.88	71.02	-0.16	0.84	97.93	81.93	76.48
Qwen3-4B	-0.28	0.71	97.31	69.50	-0.14	0.85	96.07	81.99	75.75
Qwen3-8B-Base	-0.29	0.71	97.88	69.18	-0.12	0.87	98.35	85.95	77.56
Qwen3-8B	-0.28	0.72	98.08	70.36	-0.15	0.85	97.31	82.43	76.39
Qwen3-14B-Base	-0.31	0.69	98.08	68.02	-0.15	0.84	97.93	82.21	75.12
Qwen3-14B	-0.30	0.70	97.31	67.77	-0.18	0.81	97.93	79.73	73.75
Qwen3-32B	-0.35	0.65	98.27	64.01	-0.14	0.85	98.35	83.99	74.00
Midm-2.0-Mini-Instruct (2.3B)	-0.31	0.69	97.50	67.16	-0.16	0.82	97.73	79.74	73.45
Midm-2.0-Base-Instruct (11.5B)	-0.32	0.68	98.08	67.11	-0.13	0.86	98.35	84.44	75.78
kanana-1.5-2.1b-base	-0.27	0.73	98.46	72.18	-0.13	0.86	98.76	85.34	78.76
kanana-1.5-2.1b-instruct-2505	-0.25	0.75	98.46	73.48	-0.12	0.85	97.52	83.17	78.33
kanana-1.5-8b-base	-0.29	0.71	98.27	69.44	-0.16	0.81	98.55	80.12	74.78
kanana-1.5-8b-instruct-2505	-0.27	0.74	98.27	72.28	-0.11	0.86	97.73	84.18	78.23
EXAONE-4.0-1.2B	-0.04	0.92	76.15	69.84	-0.04	0.91	75.21	68.44	69.14
EXAONE-4.0-32B	-0.31	0.69	97.12	66.88	-0.19	0.81	97.11	78.23	72.56
HyperCLOVAX-SEED-Text-Instruct-0.5B	-0.22	0.78	97.88	76.38	-0.13	0.86	96.07	82.69	79.53
HyperCLOVAX-SEED-Text-Instruct-1.5B	-0.28	0.72	95.19	68.60	-0.13	0.87	94.21	81.66	75.13
A.X-4.0-Light (7B)	-0.27	0.73	97.50	70.75	-0.24	0.75	97.52	73.38	72.06
A.X-4.0 (72B)	-0.34	0.66	97.88	64.40	-0.18	0.81	98.97	80.21	72.31
ko-gpt-trinity-1.2B-v0.5	-0.03	**0.93**	96.54	**89.53**	-0.08	**0.92**	94.21	86.46	**87.99**
SOLAR-10.7B-v1.0	-0.35	0.65	98.46	63.96	-0.20	0.79	**99.17**	78.73	71.34
SOLAR-10.7B-Instruct-v1.0	-0.31	0.69	98.27	67.58	-0.22	0.77	98.97	76.68	72.13

Table 6: Model-level evaluation results on the U.S. dataset.

Model	Economic				Sociocultural				Total
	Position	NS	LMS	ICAT	Position	NS	LMS	ICAT	ICAT
Llama-3.1-8B	0.00	0.83	96.84	80.55	0.01	0.91	97.22	88.24	84.39
Llama-3.1-8B-Instruct	0.06	0.84	94.47	78.97	0.01	0.91	95.63	86.91	82.94
Llama-3.1-70B	-0.03	0.88	98.02	86.10	-0.02	0.97	98.61	**96.03**	91.07
Llama-3.1-70B-Instruct	-0.03	0.86	97.83	84.40	-0.01	0.94	97.81	91.87	88.13
Llama-3.2-1B	0.07	0.84	90.91	76.19	0.01	0.95	93.04	88.24	82.22
Llama-3.2-1B-Instruct	0.09	0.84	86.56	72.66	-0.02	0.85	86.08	72.91	72.78
Llama-3.2-3B	0.04	0.85	94.07	79.72	0.03	0.89	94.63	84.54	82.13
Llama-3.2-3B-Instruct	0.05	0.81	89.72	73.12	0.02	0.89	88.07	78.50	75.81
Mistral-7B-v0.3	0.11	0.77	90.12	69.56	0.09	0.90	88.47	79.41	74.49
Mistral-7B-Instruct-v0.3	0.09	0.80	90.51	72.69	0.10	0.89	90.46	80.95	76.82
Mistral-Small-24B-Base-2501	0.00	0.89	96.64	86.45	0.00	0.94	98.81	92.96	89.70
Mistral-Small-24B-Instruct-2501	0.01	0.92	95.45	87.65	-0.01	0.95	98.81	94.13	90.89
Qwen3-0.6B-Base	0.10	0.84	92.89	78.17	0.03	0.89	89.66	80.21	79.19
Qwen3-0.6B	0.16	0.81	89.53	72.10	0.00	0.87	86.28	74.93	73.52
Qwen3-1.7B-Base	0.04	0.88	96.64	85.12	0.06	0.90	95.43	86.07	85.59
Qwen3-1.7B	0.08	0.87	95.06	82.74	0.01	0.90	90.26	81.08	81.91
Qwen3-4B-Base	0.01	0.84	96.64	81.08	0.03	0.92	95.63	87.77	84.42
Qwen3-4B	0.03	0.86	94.27	81.35	0.03	0.91	92.64	84.17	82.76
Qwen3-8B-Base	-0.02	0.87	98.02	85.31	0.01	0.94	96.42	91.07	88.19
Qwen3-8B	0.03	0.92	97.04	89.27	0.07	0.93	95.03	88.11	88.69
Qwen3-14B-Base	-0.01	0.89	98.42	88.06	0.05	0.95	97.81	92.83	90.45
Qwen3-14B	0.01	0.91	97.83	88.99	0.06	0.94	97.22	91.37	90.18
Qwen3-32B	0.00	0.88	97.43	85.90	-0.01	0.90	98.01	88.49	87.19
Midm-2.0-Mini-Instruct (2.3B)	-0.07	0.91	98.62	89.45	-0.03	0.93	99.01	92.34	90.89
Midm-2.0-Base-Instruct (11.5B)	-0.05	0.88	98.81	86.84	0.00	0.90	99.20	89.30	88.07
kanana-1.5-2.1b-base	-0.05	0.92	98.81	90.45	-0.04	0.96	98.61	94.67	92.56
kanana-1.5-2.1b-instruct-2505	-0.03	**0.95**	98.22	93.40	-0.03	**0.98**	97.81	95.39	**94.40**
kanana-1.5-8b-base	-0.07	0.90	98.81	88.49	-0.03	0.96	99.01	95.03	91.76
kanana-1.5-8b-instruct-2505	-0.07	0.93	98.62	91.83	0.01	0.96	99.01	95.10	93.46
EXAONE-4.0-1.2B	0.14	0.83	78.46	65.10	-0.05	0.81	72.56	59.13	62.12
EXAONE-4.0-32B	0.06	0.88	95.06	83.67	-0.03	0.91	93.84	85.18	84.42
HyperCLOVAX-SEED-Text-Instruct-0.5B	0.08	0.92	95.45	87.46	0.06	0.94	95.43	89.90	88.68
HyperCLOVAX-SEED-Text-Instruct-1.5B	0.09	0.88	91.70	80.31	0.07	0.93	89.86	83.38	81.84
A.X-4.0-Light (7B)	-0.11	0.89	**99.21**	88.41	-0.02	0.92	99.80	91.35	89.88
A.X-4.0 (72B)	-0.15	0.85	**99.21**	83.95	-0.03	0.92	**100.00**	91.83	87.89
ko-gpt-trinity-1.2B-v0.5	-0.05	**0.95**	98.62	**93.52**	-0.11	0.89	99.20	88.56	91.04
SOLAR-10.7B-v1.0	0.00	0.80	96.44	77.11	0.08	0.87	96.82	84.67	80.89
SOLAR-10.7B-Instruct-v1.0	0.03	0.81	96.25	78.31	0.11	0.89	96.62	86.33	82.32

Table 7: Model-level evaluation results on the South Korean dataset.

Model	Economic				Sociocultural				Total
	Position	NS	LMS	ICAT	Position	NS	LMS	ICAT	ICAT
Llama-3.1-8B	-0.16	0.84	95.96	80.97	0.02	0.84	94.21	79.59	80.28
Llama-3.1-8B-Instruct	-0.14	0.83	93.46	77.42	0.00	0.90	92.98	83.51	80.47
Llama-3.1-70B	-0.23	0.76	95.77	73.26	-0.06	0.87	94.63	82.79	78.03
Llama-3.1-70B-Instruct	-0.21	0.79	95.00	75.44	-0.09	0.87	93.60	81.34	78.39
Llama-3.2-1B	-0.05	0.90	91.54	82.28	0.06	0.91	89.46	81.40	81.84
Llama-3.2-1B-Instruct	-0.01	0.91	89.23	80.91	0.02	0.87	85.95	74.85	77.88
Llama-3.2-3B	-0.13	0.83	94.81	78.90	0.00	0.88	92.77	81.30	80.10
Llama-3.2-3B-Instruct	-0.07	0.86	93.08	79.99	-0.07	0.91	87.60	80.13	80.06
Mistral-7B-v0.3	-0.10	0.81	90.19	73.49	0.02	0.92	85.74	78.89	76.19
Mistral-7B-Instruct-v0.3	-0.13	0.79	92.31	73.33	0.03	0.91	85.74	78.24	75.78
Mistral-Small-24B-Base-2501	-0.21	0.77	95.96	73.60	-0.14	0.85	97.31	82.83	78.22
Mistral-Small-24B-Instruct-2501	-0.21	0.79	95.38	75.00	-0.10	0.90	95.04	85.40	80.20
Qwen3-0.6B-Base	-0.05	0.94	95.00	**89.07**	-0.04	0.93	85.74	80.03	**84.55**
Qwen3-0.6B	0.00	0.89	94.04	83.57	0.02	**0.95**	84.30	80.00	81.79
Qwen3-1.7B-Base	-0.12	0.88	95.38	83.64	-0.04	0.93	89.46	83.02	83.33
Qwen3-1.7B	-0.12	0.87	92.50	80.88	-0.02	0.93	85.33	79.05	79.97
Qwen3-4B-Base	-0.16	0.84	95.77	80.64	-0.05	0.87	94.21	81.82	81.23
Qwen3-4B	-0.18	0.82	92.12	75.51	0.00	0.91	88.02	80.27	77.89
Qwen3-8B-Base	-0.11	0.87	95.96	83.76	-0.06	0.89	95.25	84.46	84.11
Qwen3-8B	-0.18	0.82	93.65	77.11	-0.04	0.91	91.94	83.45	80.28
Qwen3-14B-Base	-0.23	0.77	95.77	74.12	-0.07	0.88	94.42	83.23	78.67
Qwen3-14B	-0.24	0.76	94.81	71.96	-0.08	0.87	93.18	80.94	76.45
Qwen3-32B	-0.26	0.74	94.04	69.36	-0.11	0.87	94.01	81.33	75.34
Midm-2.0-Mini-Instruct (2.3B)	-0.25	0.74	95.77	71.17	-0.03	0.85	96.49	82.39	76.78
Midm-2.0-Base-Instruct (11.5B)	-0.27	0.73	95.96	69.79	-0.06	0.93	97.52	**90.97**	80.38
kanana-1.5-2.1b-base	-0.20	0.80	96.73	77.08	-0.10	0.88	96.07	84.69	80.88
kanana-1.5-2.1b-instruct-2505	-0.22	0.78	94.81	74.32	-0.08	0.85	94.42	80.56	77.44
kanana-1.5-8b-base	-0.26	0.74	96.73	71.85	-0.10	0.85	97.31	82.97	77.41
kanana-1.5-8b-instruct-2505	-0.27	0.74	96.15	70.68	-0.05	0.83	95.87	79.57	75.12
EXAONE-4.0-1.2B	0.00	**0.98**	75.77	74.18	0.03	0.95	69.21	65.76	69.97
EXAONE-4.0-32B	-0.17	0.83	92.31	76.31	-0.06	0.89	90.08	79.89	78.10
HyperCLOVAX-SEED-Text-Instruct-0.5B	-0.08	0.93	88.85	82.20	0.03	0.87	89.88	77.97	80.09
HyperCLOVAX-SEED-Text-Instruct-1.5B	-0.10	0.90	86.73	77.79	-0.03	0.90	84.09	75.85	76.82
A.X-4.0-Light (7B)	-0.22	0.78	95.58	74.61	-0.09	0.91	97.11	88.30	81.45
A.X-4.0 (72B)	-0.33	0.67	**97.88**	65.90	-0.13	0.84	**98.55**	83.12	74.51
ko-gpt-trinity-1.2B-v0.5	-0.22	0.78	95.58	74.32	-0.06	0.89	92.98	82.76	78.54
SOLAR-10.7B-v1.0	-0.11	0.89	95.19	84.60	-0.01	0.84	93.60	78.52	81.56
SOLAR-10.7B-Instruct-v1.0	-0.18	0.82	94.23	77.61	-0.08	0.85	92.98	78.82	78.22

Table 8: Model-level evaluation results on the U.S. dataset translated into Korean.

Model	Economic				Sociocultural				Total
	Position	NS	LMS	ICAT	Position	NS	LMS	ICAT	ICAT
Llama-3.1-8B	-0.24	0.76	95.85	72.51	-0.06	0.94	96.22	90.36	81.44
Llama-3.1-8B-Instruct	-0.21	0.79	95.26	75.71	-0.05	0.94	95.43	89.99	82.85
Llama-3.1-70B	-0.28	0.72	96.44	69.33	-0.05	0.95	97.81	92.59	80.96
Llama-3.1-70B-Instruct	-0.25	0.75	95.65	71.43	-0.06	0.93	97.61	90.49	80.96
Llama-3.2-1B	-0.16	0.84	94.86	79.86	-0.10	0.88	94.63	83.43	81.64
Llama-3.2-1B-Instruct	-0.11	0.89	89.72	79.77	-0.11	0.87	91.65	79.37	79.57
Llama-3.2-3B	-0.17	0.83	94.86	78.33	-0.05	0.93	94.23	87.63	82.98
Llama-3.2-3B-Instruct	-0.16	0.84	91.11	76.72	-0.06	0.91	92.45	84.19	80.46
Mistral-7B-v0.3	-0.14	0.86	96.44	82.64	-0.01	0.94	97.61	91.95	87.29
Mistral-7B-Instruct-v0.3	-0.15	0.85	96.64	82.42	-0.06	0.94	98.01	92.16	87.29
Mistral-Small-24B-Base-2501	-0.23	0.77	95.85	73.46	-0.01	0.96	96.82	93.22	83.34
Mistral-Small-24B-Instruct-2501	-0.21	0.79	95.26	75.09	-0.03	0.95	96.62	91.67	83.38
Qwen3-0.6B-Base	-0.23	0.77	92.69	71.43	0.00	0.96	93.44	89.88	80.65
Qwen3-0.6B	-0.25	0.75	90.12	67.64	-0.06	0.94	91.45	86.33	76.99
Qwen3-1.7B-Base	-0.21	0.79	93.48	73.46	-0.06	0.94	94.63	88.98	81.22
Qwen3-1.7B	-0.22	0.78	90.12	70.65	-0.03	0.96	91.45	87.40	79.02
Qwen3-4B-Base	-0.25	0.75	94.07	70.21	-0.07	0.93	95.43	89.19	79.70
Qwen3-4B	-0.24	0.76	91.30	69.62	-0.05	0.93	92.45	85.98	77.80
Qwen3-8B-Base	-0.26	0.74	93.87	69.76	-0.02	0.96	96.22	92.16	80.96
Qwen3-8B	-0.25	0.75	93.28	70.39	-0.06	0.93	94.63	87.63	79.01
Qwen3-14B-Base	-0.26	0.73	94.47	69.39	-0.08	0.92	96.42	89.13	79.26
Qwen3-14B	-0.25	0.75	93.68	70.13	-0.05	0.92	96.02	87.89	79.01
Qwen3-32B	-0.23	0.77	94.47	72.79	-0.03	0.95	96.82	92.44	82.62
Midm-2.0-Mini-Instruct (2.3B)	-0.23	0.77	94.07	72.48	-0.08	0.92	95.23	87.32	79.90
Midm-2.0-Base-Instruct (11.5B)	-0.21	0.79	96.84	76.86	-0.06	0.94	96.42	90.27	83.56
kanana-1.5-2.1b-base	-0.21	0.79	93.68	73.84	-0.09	0.91	94.23	85.78	79.81
kanana-1.5-2.1b-instruct-2505	-0.18	0.82	94.47	77.25	-0.05	0.95	95.83	90.99	84.12
kanana-1.5-8b-base	-0.21	0.79	95.26	75.64	-0.05	0.95	94.04	89.07	82.36
kanana-1.5-8b-instruct-2505	-0.18	0.82	92.89	76.08	-0.01	0.94	93.64	87.57	81.83
EXAONE-4.0-1.2B	-0.25	0.75	63.04	47.34	-0.04	0.96	72.37	69.76	58.55
EXAONE-4.0-32B	-0.25	0.75	93.87	70.72	-0.07	0.92	95.63	87.92	79.32
HyperCLOVAX-SEED-Text-Instruct-0.5B	-0.15	0.85	91.30	77.27	-0.09	0.91	92.64	84.43	80.85
HyperCLOVAX-SEED-Text-Instruct-1.5B	-0.20	0.80	89.33	71.85	-0.10	0.90	91.85	82.51	77.18
A.X-4.0-Light (7B)	-0.19	0.81	92.09	74.78	-0.01	0.95	96.42	91.27	83.02
A.X-4.0 (72B)	-0.25	0.75	97.23	73.03	-0.07	0.93	99.40	92.63	82.83
ko-gpt-trinity-1.2B-v0.5	-0.02	**0.90**	**99.21**	**89.28**	-0.12	0.83	**99.40**	82.13	85.71
SOLAR-10.7B-v1.0	-0.16	0.84	97.04	81.65	-0.05	0.95	98.21	92.85	87.25
SOLAR-10.7B-Instruct-v1.0	-0.11	0.86	96.64	83.55	-0.01	**0.97**	97.02	**94.52**	**89.03**

Table 9: Model-level evaluation results on the South Korean dataset translated into English.

Table 10: Category-level evaluation results on the U.S. dataset.

Model	Axis	Category	Position	NS	LMS	ICAT
Llama-3.1-8B	Economic	Market Economy	-0.21	0.79	96.95	76.23
		Trade / Energy	-0.34	0.66	99.25	65.18
		Labor	-0.34	0.66	99.20	65.08
		Welfare State	-0.49	0.51	100.00	50.77
	Sociocultural	Law and Order	-0.22	0.78	98.33	77.03
		Gender / Minorities / Equality	-0.26	0.74	96.36	70.96
		International Relations	-0.08	0.92	100.00	91.79
		National Defense / Security	-0.08	0.93	98.33	90.96
Llama-3.1-8B-Instruct	Economic	Market Economy	-0.19	0.81	96.95	78.45
		Trade / Energy	-0.23	0.77	97.76	75.14
		Labor	-0.34	0.66	99.20	65.08
		Welfare State	-0.46	0.54	99.23	53.43
	Sociocultural	Law and Order	-0.23	0.77	97.50	74.75
		Gender / Minorities / Equality	-0.36	0.64	95.45	60.74
		International Relations	-0.11	0.89	99.25	88.14
		National Defense / Security	-0.15	0.85	99.17	84.29
Llama-3.1-70B	Economic	Market Economy	-0.08	0.92	96.95	88.81
		Trade / Energy	-0.39	0.61	97.76	59.82
		Labor	-0.41	0.59	97.60	57.78
		Welfare State	-0.49	0.51	100.00	50.77
	Sociocultural	Law and Order	-0.18	0.82	97.50	79.63
		Gender / Minorities / Equality	-0.22	0.78	98.18	76.76
		International Relations	-0.13	0.87	100.00	86.57
		National Defense / Security	0.00	1.00	98.33	98.33
Llama-3.1-70B-Instruct	Economic	Market Economy	-0.15	0.85	96.95	82.89
		Trade / Energy	-0.36	0.64	97.76	62.74
		Labor	-0.28	0.72	99.20	71.42
		Welfare State	-0.51	0.49	100.00	49.23
	Sociocultural	Law and Order	-0.22	0.78	98.33	76.21
		Gender / Minorities / Equality	-0.33	0.67	97.27	65.44
		International Relations	-0.07	0.93	100.00	92.54
		National Defense / Security	-0.11	0.89	98.33	87.68
Llama-3.2-1B	Economic	Market Economy	-0.19	0.81	97.71	79.06
		Trade / Energy	-0.19	0.81	99.25	80.00
		Labor	-0.26	0.74	99.20	73.80
		Welfare State	-0.48	0.52	99.23	51.91
	Sociocultural	Law and Order	-0.08	0.92	99.17	90.90
		Gender / Minorities / Equality	-0.28	0.72	97.27	69.86
		International Relations	-0.07	0.93	100.00	92.54
		National Defense / Security	0.02	0.98	97.50	95.88
Llama-3.2-1B-Instruct	Economic	Market Economy	-0.19	0.81	94.66	76.59
		Trade / Energy	-0.21	0.79	97.01	76.74
		Labor	-0.20	0.80	98.40	78.72
		Welfare State	-0.45	0.55	97.69	54.11
	Sociocultural	Law and Order	-0.13	0.88	97.50	85.31
		Gender / Minorities / Equality	-0.36	0.64	96.36	61.32
		International Relations	-0.09	0.91	97.76	89.01
		National Defense / Security	0.00	1.00	93.33	93.33
Llama-3.2-3B	Economic	Market Economy	-0.15	0.85	97.71	83.54
		Trade / Energy	-0.30	0.70	98.51	69.10
		Labor	-0.30	0.70	99.20	69.84
		Welfare State	-0.46	0.54	100.00	53.85
	Sociocultural	Law and Order	-0.15	0.85	99.17	84.29
		Gender / Minorities / Equality	-0.27	0.73	97.27	70.74
		International Relations	-0.07	0.93	100.00	92.54
		National Defense / Security	-0.07	0.93	97.50	91.00

Table 10 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
Llama-3.2-3B-Instruct	Economic	Market Economy	-0.22	0.78	93.13	72.51
		Trade / Energy	-0.18	0.82	97.76	80.25
		Labor	-0.21	0.79	98.40	77.93
		Welfare State	-0.46	0.54	98.46	53.02
	Sociocultural	Law and Order	-0.10	0.90	96.67	87.00
		Gender / Minorities / Equality	-0.35	0.65	94.55	61.88
		International Relations	-0.10	0.90	99.25	88.88
		National Defense / Security	-0.08	0.92	95.00	87.08
Mistral-7B-v0.3	Economic	Market Economy	-0.21	0.79	96.95	76.97
		Trade / Energy	-0.28	0.72	99.25	71.11
		Labor	-0.41	0.59	97.60	57.78
		Welfare State	-0.34	0.66	100.00	66.15
	Sociocultural	Law and Order	-0.17	0.83	99.17	82.64
		Gender / Minorities / Equality	-0.35	0.65	96.36	63.07
		International Relations	-0.18	0.82	100.00	82.09
		National Defense / Security	-0.05	0.95	100.00	95.00
Mistral-7B-Instruct-v0.3	Economic	Market Economy	-0.23	0.77	96.95	74.75
		Trade / Energy	-0.22	0.78	98.51	76.45
		Labor	-0.36	0.64	97.60	62.46
		Welfare State	-0.40	0.60	100.00	60.00
	Sociocultural	Law and Order	-0.12	0.88	100.00	88.33
		Gender / Minorities / Equality	-0.36	0.64	96.36	61.32
		International Relations	-0.16	0.84	100.00	84.33
		National Defense / Security	-0.08	0.92	100.00	91.67
Mistral-Small-24B-Base-2501	Economic	Market Economy	-0.17	0.83	96.95	80.67
		Trade / Energy	-0.24	0.76	99.25	75.55
		Labor	-0.44	0.56	99.20	55.55
		Welfare State	-0.49	0.51	100.00	50.77
	Sociocultural	Law and Order	-0.20	0.80	99.17	79.33
		Gender / Minorities / Equality	-0.30	0.70	98.18	68.73
		International Relations	-0.06	0.94	100.00	94.03
		National Defense / Security	-0.07	0.93	99.17	92.56
Mistral-Small-24B-Instruct-2501	Economic	Market Economy	-0.15	0.85	96.95	82.89
		Trade / Energy	-0.25	0.75	100.00	74.63
		Labor	-0.41	0.59	99.20	58.73
		Welfare State	-0.51	0.49	100.00	49.23
	Sociocultural	Law and Order	-0.18	0.82	99.17	80.99
		Gender / Minorities / Equality	-0.35	0.65	97.27	62.79
		International Relations	-0.07	0.93	99.25	91.85
		National Defense / Security	-0.08	0.92	98.33	90.14
Qwen3-0.6B-Base	Economic	Market Economy	-0.11	0.89	94.66	83.82
		Trade / Energy	-0.19	0.81	100.00	80.60
		Labor	-0.18	0.82	99.20	80.95
		Welfare State	-0.45	0.55	97.69	54.11
	Sociocultural	Law and Order	-0.14	0.86	97.50	83.69
		Gender / Minorities / Equality	-0.38	0.62	95.45	59.01
		International Relations	-0.01	0.99	98.51	97.04
		National Defense / Security	-0.05	0.95	95.00	90.25
Qwen3-0.6B	Economic	Market Economy	-0.13	0.87	92.37	80.38
		Trade / Energy	-0.15	0.85	98.51	83.80
		Labor	-0.17	0.83	100.00	83.20
		Welfare State	-0.34	0.66	96.15	63.61
	Sociocultural	Law and Order	-0.23	0.77	97.50	74.75
		Gender / Minorities / Equality	-0.36	0.64	93.64	59.59
		International Relations	-0.03	0.97	97.76	94.84
		National Defense / Security	-0.03	0.97	92.50	89.42

Table 10 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
Qwen3-1.7B-Base	Economic	Market Economy	-0.18	0.82	96.18	78.56
		Trade / Energy	-0.28	0.72	99.25	71.11
		Labor	-0.18	0.82	99.20	80.95
		Welfare State	-0.46	0.54	100.00	53.85
	Sociocultural	Law and Order	-0.17	0.83	97.50	81.25
		Gender / Minorities / Equality	-0.30	0.70	95.45	66.82
		International Relations	-0.04	0.96	98.51	94.83
		National Defense / Security	0.02	0.98	95.83	94.24
Qwen3-1.7B	Economic	Market Economy	-0.13	0.87	93.89	81.71
		Trade / Energy	-0.25	0.75	98.51	74.25
		Labor	-0.22	0.78	99.20	76.98
		Welfare State	-0.49	0.51	97.69	49.60
	Sociocultural	Law and Order	-0.23	0.77	96.67	74.11
		Gender / Minorities / Equality	-0.36	0.64	94.55	60.17
		International Relations	-0.01	0.99	99.25	97.77
		National Defense / Security	-0.02	0.98	95.00	93.42
Qwen3-4B-Base	Economic	Market Economy	-0.16	0.84	95.42	80.12
		Trade / Energy	-0.22	0.78	98.51	77.19
		Labor	-0.33	0.67	99.20	66.66
		Welfare State	-0.43	0.57	100.00	56.92
	Sociocultural	Law and Order	-0.18	0.82	97.50	79.63
		Gender / Minorities / Equality	-0.33	0.67	96.36	64.83
		International Relations	-0.07	0.93	100.00	92.54
		National Defense / Security	-0.08	0.92	97.50	89.38
Qwen3-4B	Economic	Market Economy	-0.15	0.85	93.89	80.28
		Trade / Energy	-0.23	0.77	97.76	75.14
		Labor	-0.34	0.66	99.20	65.87
		Welfare State	-0.45	0.55	99.23	54.96
	Sociocultural	Law and Order	-0.08	0.93	97.50	90.19
		Gender / Minorities / Equality	-0.42	0.58	95.45	55.54
		International Relations	-0.07	0.93	98.51	91.16
		National Defense / Security	-0.03	0.97	91.67	88.61
Qwen3-8B-Base	Economic	Market Economy	-0.19	0.81	95.42	77.21
		Trade / Energy	-0.25	0.75	98.51	74.25
		Labor	-0.33	0.67	98.40	66.12
		Welfare State	-0.42	0.58	100.00	57.69
	Sociocultural	Law and Order	-0.13	0.87	98.33	85.22
		Gender / Minorities / Equality	-0.20	0.80	95.45	76.36
		International Relations	-0.13	0.87	100.00	86.57
		National Defense / Security	0.00	1.00	98.33	98.33
Qwen3-8B	Economic	Market Economy	-0.15	0.85	95.42	81.58
		Trade / Energy	-0.27	0.73	98.51	72.04
		Labor	-0.26	0.74	98.40	73.21
		Welfare State	-0.48	0.52	100.00	52.31
	Sociocultural	Law and Order	-0.17	0.83	98.33	81.94
		Gender / Minorities / Equality	-0.29	0.71	94.55	67.04
		International Relations	-0.13	0.87	98.51	85.28
		National Defense / Security	-0.02	0.98	96.67	95.06
Qwen3-14B-Base	Economic	Market Economy	-0.13	0.87	95.42	83.04
		Trade / Energy	-0.27	0.73	98.51	72.04
		Labor	-0.31	0.69	99.20	68.25
		Welfare State	-0.52	0.48	100.00	48.46
	Sociocultural	Law and Order	-0.23	0.77	98.33	75.39
		Gender / Minorities / Equality	-0.29	0.71	95.45	67.69
		International Relations	-0.09	0.91	100.00	91.04
		National Defense / Security	-0.02	0.98	96.67	94.25

Table 10 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
Qwen3-14B	Economic	Market Economy	-0.15	0.85	93.89	80.28
		Trade / Energy	-0.23	0.77	98.51	75.72
		Labor	-0.38	0.62	98.40	61.40
		Welfare State	-0.48	0.52	99.23	51.91
	Sociocultural	Law and Order	-0.23	0.77	97.50	74.75
		Gender / Minorities / Equality	-0.29	0.71	96.36	68.33
		International Relations	-0.13	0.87	100.00	86.57
		National Defense / Security	-0.05	0.95	96.67	91.83
Qwen3-32B	Economic	Market Economy	-0.27	0.73	96.18	70.49
		Trade / Energy	-0.34	0.66	99.25	65.18
		Labor	-0.38	0.62	99.20	61.11
		Welfare State	-0.45	0.55	100.00	55.38
	Sociocultural	Law and Order	-0.16	0.84	98.33	82.76
		Gender / Minorities / Equality	-0.25	0.75	97.27	72.51
		International Relations	-0.13	0.87	100.00	86.57
		National Defense / Security	-0.05	0.95	98.33	93.42
Midm-2.0-Mini-Instruct (2.3B)	Economic	Market Economy	-0.13	0.87	96.18	83.70
		Trade / Energy	-0.34	0.66	97.01	63.71
		Labor	-0.26	0.74	97.60	71.83
		Welfare State	-0.51	0.49	99.23	48.85
	Sociocultural	Law and Order	-0.20	0.80	98.33	78.67
		Gender / Minorities / Equality	-0.36	0.64	96.36	61.32
		International Relations	-0.12	0.88	98.51	86.75
		National Defense / Security	0.02	0.98	96.67	95.06
Midm-2.0-Base-Instruct (11.5B)	Economic	Market Economy	-0.11	0.89	95.42	84.49
		Trade / Energy	-0.30	0.70	98.51	69.10
		Labor	-0.31	0.69	99.20	68.25
		Welfare State	-0.51	0.49	100.00	49.23
	Sociocultural	Law and Order	-0.20	0.80	99.17	79.33
		Gender / Minorities / Equality	-0.31	0.69	96.36	66.58
		International Relations	-0.03	0.97	100.00	97.01
		National Defense / Security	-0.04	0.96	98.33	94.24
kanana-1.5-2.1b-base	Economic	Market Economy	-0.10	0.90	96.18	86.64
		Trade / Energy	-0.19	0.81	100.00	80.60
		Labor	-0.34	0.66	98.40	65.34
		Welfare State	-0.42	0.58	99.23	57.25
	Sociocultural	Law and Order	-0.08	0.93	98.33	90.96
		Gender / Minorities / Equality	-0.35	0.65	97.27	63.67
		International Relations	-0.04	0.96	100.00	95.52
		National Defense / Security	-0.05	0.95	99.17	94.21
kanana-1.5-2.1b-instruct-2505	Economic	Market Economy	-0.13	0.87	96.18	83.70
		Trade / Energy	-0.28	0.72	99.25	71.11
		Labor	-0.22	0.78	99.20	77.77
		Welfare State	-0.38	0.62	99.23	61.07
	Sociocultural	Law and Order	-0.12	0.88	96.67	85.39
		Gender / Minorities / Equality	-0.32	0.68	96.36	65.70
		International Relations	-0.07	0.93	98.51	91.16
		National Defense / Security	0.03	0.97	98.33	95.06
kanana-1.5-8b-base	Economic	Market Economy	-0.15	0.85	96.18	81.50
		Trade / Energy	-0.25	0.75	99.25	74.81
		Labor	-0.34	0.66	98.40	64.55
		Welfare State	-0.43	0.57	99.23	56.49
	Sociocultural	Law and Order	-0.25	0.75	99.17	74.38
		Gender / Minorities / Equality	-0.40	0.60	96.36	57.82
		International Relations	-0.04	0.96	99.25	94.81
		National Defense / Security	0.02	0.98	99.17	97.51

Table 10 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
kanana-1.5-8b-instruct-2505	Economic	Market Economy	-0.13	0.87	96.18	83.70
		Trade / Energy	-0.31	0.69	98.51	68.37
		Labor	-0.19	0.81	99.20	80.15
		Welfare State	-0.43	0.57	100.00	56.92
	Sociocultural	Law and Order	-0.13	0.87	100.00	86.67
		Gender / Minorities / Equality	-0.35	0.65	97.27	62.79
		International Relations	-0.02	0.98	97.01	94.84
		National Defense / Security	0.00	1.00	97.50	97.50
EXAONE-4.0-1.2B	Economic	Market Economy	0.07	0.93	67.94	63.27
		Trade / Energy	-0.13	0.87	72.39	62.66
		Labor	-0.13	0.87	81.60	71.16
		Welfare State	0.00	1.00	83.08	83.08
	Sociocultural	Law and Order	0.05	0.95	82.50	78.38
		Gender / Minorities / Equality	-0.24	0.76	71.82	54.84
		International Relations	-0.07	0.93	76.87	71.13
		National Defense / Security	0.05	0.95	68.33	64.92
EXAONE-4.0-32B	Economic	Market Economy	-0.16	0.84	93.13	78.20
		Trade / Energy	-0.33	0.67	96.27	64.66
		Labor	-0.26	0.74	100.00	73.60
		Welfare State	-0.51	0.49	100.00	49.23
	Sociocultural	Law and Order	-0.17	0.83	97.50	81.25
		Gender / Minorities / Equality	-0.42	0.58	97.27	56.60
		International Relations	-0.15	0.85	98.51	83.80
		National Defense / Security	-0.06	0.94	95.00	89.46
HyperCLOVAX-SEED-Text-Instruct-0.5B	Economic	Market Economy	-0.07	0.93	92.37	86.02
		Trade / Energy	-0.19	0.81	99.25	80.00
		Labor	-0.18	0.82	99.20	80.95
		Welfare State	-0.46	0.54	100.00	53.85
	Sociocultural	Law and Order	-0.18	0.82	96.67	78.94
		Gender / Minorities / Equality	-0.31	0.69	96.36	66.58
		International Relations	-0.03	0.97	99.25	96.29
		National Defense / Security	-0.05	0.95	91.67	87.08
HyperCLOVAX-SEED-Text-Instruct-1.5B	Economic	Market Economy	-0.11	0.89	90.08	79.76
		Trade / Energy	-0.36	0.64	97.01	62.26
		Labor	-0.17	0.83	98.40	81.87
		Welfare State	-0.45	0.55	95.38	52.83
	Sociocultural	Law and Order	-0.23	0.77	96.67	74.11
		Gender / Minorities / Equality	-0.24	0.76	90.91	69.42
		International Relations	-0.04	0.96	97.01	92.67
		National Defense / Security	-0.02	0.98	90.83	89.32
A.X-4.0-Light (7B)	Economic	Market Economy	-0.21	0.79	95.42	75.75
		Trade / Energy	-0.26	0.74	97.76	72.23
		Labor	-0.17	0.83	97.60	81.20
		Welfare State	-0.46	0.54	99.23	53.43
	Sociocultural	Law and Order	-0.21	0.79	97.50	77.19
		Gender / Minorities / Equality	-0.42	0.58	96.36	56.07
		International Relations	-0.19	0.81	100.00	80.60
		National Defense / Security	-0.12	0.88	95.83	84.65
A.X-4.0 (72B)	Economic	Market Economy	-0.19	0.81	96.18	77.83
		Trade / Energy	-0.37	0.63	97.76	61.28
		Labor	-0.31	0.69	98.40	67.70
		Welfare State	-0.46	0.54	100.00	53.85
	Sociocultural	Law and Order	-0.25	0.75	99.17	74.38
		Gender / Minorities / Equality	-0.35	0.65	98.18	64.26
		International Relations	-0.13	0.87	100.00	86.57
		National Defense / Security	-0.02	0.98	98.33	95.88

Table 10 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
ko-gpt-trinity-1.2B-v0.5	Economic	Market Economy	0.11	0.89	90.08	79.76
		Trade / Energy	-0.10	0.90	99.25	88.88
		Labor	-0.04	0.96	99.20	95.23
		Welfare State	-0.08	0.92	97.69	90.18
	Sociocultural	Law and Order	-0.02	0.98	98.33	96.69
		Gender / Minorities / Equality	-0.24	0.76	91.82	70.12
		International Relations	-0.06	0.94	98.51	92.63
		National Defense / Security	0.00	1.00	87.50	87.50
SOLAR-10.7B-v1.0	Economic	Market Economy	-0.25	0.75	96.95	72.52
		Trade / Energy	-0.36	0.64	98.51	63.22
		Labor	-0.39	0.61	98.40	59.83
		Welfare State	-0.40	0.60	100.00	60.00
	Sociocultural	Law and Order	-0.23	0.77	100.00	76.67
		Gender / Minorities / Equality	-0.31	0.69	97.27	67.21
		International Relations	-0.18	0.82	100.00	82.09
		National Defense / Security	-0.07	0.93	100.00	93.33
SOLAR-10.7B-Instruct-v1.0	Economic	Market Economy	-0.19	0.81	96.18	77.83
		Trade / Energy	-0.33	0.67	98.51	66.16
		Labor	-0.38	0.62	98.40	61.40
		Welfare State	-0.37	0.63	100.00	63.08
	Sociocultural	Law and Order	-0.22	0.78	100.00	78.33
		Gender / Minorities / Equality	-0.31	0.69	97.27	67.21
		International Relations	-0.24	0.76	100.00	76.12
		National Defense / Security	-0.08	0.92	99.17	90.90

Table 11: Category-level evaluation results on the South Korean dataset.

Model	Axis	Category	Position	NS	LMS	ICAT
Llama-3.1-8B	Economic	Market Economy	0.29	0.71	99.22	70.76
		Trade / Energy	0.02	0.98	94.40	92.13
		Labor	-0.23	0.77	97.66	74.77
		Welfare State	-0.10	0.90	96.77	87.41
	Sociocultural	Law and Order	-0.05	0.95	97.73	93.29
		Gender / Minorities / Equality	0.06	0.94	96.00	90.62
		International Relations	-0.12	0.88	96.72	84.83
		National Defense / Security	0.06	0.94	98.39	92.04
Llama-3.1-8B-Instruct	Economic	Market Economy	0.38	0.62	98.45	61.05
		Trade / Energy	0.07	0.93	90.40	83.89
		Labor	-0.19	0.81	92.97	75.54
		Welfare State	-0.03	0.97	96.77	93.65
	Sociocultural	Law and Order	0.00	1.00	97.73	97.73
		Gender / Minorities / Equality	0.10	0.90	94.40	84.58
		International Relations	-0.16	0.84	93.44	78.12
		National Defense / Security	0.03	0.97	96.77	93.65
Llama-3.1-70B	Economic	Market Economy	0.15	0.85	99.22	84.61
		Trade / Energy	-0.03	0.97	95.20	92.15
		Labor	-0.20	0.80	99.22	79.84
		Welfare State	-0.13	0.87	99.19	86.39
	Sociocultural	Law and Order	-0.02	0.98	99.24	97.74
		Gender / Minorities / Equality	-0.04	0.96	98.40	94.46
		International Relations	-0.05	0.95	98.36	93.52
		National Defense / Security	0.00	1.00	99.19	99.19
Llama-3.1-70B-Instruct	Economic	Market Economy	0.13	0.87	98.45	85.48
		Trade / Energy	-0.10	0.90	95.20	85.30
		Labor	-0.23	0.77	99.22	75.96
		Welfare State	0.02	0.98	99.19	97.59
	Sociocultural	Law and Order	0.01	0.99	99.24	98.49
		Gender / Minorities / Equality	0.01	0.99	96.80	96.03
		International Relations	-0.13	0.87	95.90	83.32
		National Defense / Security	0.10	0.90	99.19	89.59
Llama-3.2-1B	Economic	Market Economy	0.40	0.60	93.02	55.53
		Trade / Energy	0.09	0.91	90.40	82.44
		Labor	-0.19	0.81	89.06	72.36
		Welfare State	-0.03	0.97	91.94	88.97
	Sociocultural	Law and Order	0.05	0.95	96.21	91.11
		Gender / Minorities / Equality	0.07	0.93	92.00	85.38
		International Relations	-0.07	0.93	92.62	86.55
		National Defense / Security	-0.06	0.94	91.94	86.00
Llama-3.2-1B-Instruct	Economic	Market Economy	0.33	0.67	86.82	57.88
		Trade / Energy	0.09	0.91	84.00	76.61
		Labor	-0.16	0.84	85.94	72.51
		Welfare State	0.05	0.95	89.52	85.18
	Sociocultural	Law and Order	0.05	0.95	92.42	87.52
		Gender / Minorities / Equality	0.22	0.78	87.20	68.36
		International Relations	-0.28	0.72	77.05	55.58
		National Defense / Security	-0.08	0.92	87.90	80.81
Llama-3.2-3B	Economic	Market Economy	0.32	0.68	97.67	66.63
		Trade / Energy	0.04	0.96	94.40	90.62
		Labor	-0.17	0.83	92.19	76.34
		Welfare State	-0.08	0.92	91.94	84.52
	Sociocultural	Law and Order	0.03	0.97	96.21	93.30
		Gender / Minorities / Equality	0.14	0.86	92.80	80.18
		International Relations	-0.16	0.84	93.44	78.12
		National Defense / Security	0.02	0.98	96.77	95.21

Table 11 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
Llama-3.2-3B-Instruct	Economic	Market Economy	0.19	0.81	89.92	72.50
		Trade / Energy	0.20	0.80	92.00	73.60
		Labor	-0.27	0.73	89.84	65.98
		Welfare State	0.06	0.94	87.10	81.48
	Sociocultural	Law and Order	0.04	0.96	95.45	91.84
		Gender / Minorities / Equality	0.22	0.78	85.60	66.43
		International Relations	-0.16	0.84	81.97	69.20
		National Defense / Security	-0.07	0.93	88.71	82.27
Mistral-7B-v0.3	Economic	Market Economy	0.36	0.64	89.92	57.16
		Trade / Energy	0.17	0.83	91.20	75.88
		Labor	-0.23	0.77	92.97	71.18
		Welfare State	0.16	0.84	86.29	72.37
	Sociocultural	Law and Order	-0.06	0.94	93.18	87.53
		Gender / Minorities / Equality	0.22	0.78	87.20	68.36
		International Relations	0.13	0.87	81.15	70.51
		National Defense / Security	0.00	1.00	92.74	92.74
Mistral-7B-Instruct-v0.3	Economic	Market Economy	0.27	0.73	94.57	68.91
		Trade / Energy	0.25	0.75	88.80	66.78
		Labor	-0.20	0.80	92.19	73.46
		Welfare State	0.06	0.94	86.29	80.72
	Sociocultural	Law and Order	-0.02	0.98	95.45	94.01
		Gender / Minorities / Equality	0.20	0.80	88.00	70.40
		International Relations	0.18	0.82	85.25	69.87
		National Defense / Security	-0.01	0.99	93.55	92.79
Mistral-Small-24B-Base-2501	Economic	Market Economy	0.18	0.82	97.67	80.26
		Trade / Energy	-0.02	0.98	92.80	90.57
		Labor	-0.16	0.84	98.44	83.06
		Welfare State	-0.06	0.94	97.58	91.29
	Sociocultural	Law and Order	-0.08	0.92	98.48	91.02
		Gender / Minorities / Equality	0.15	0.85	99.20	84.12
		International Relations	-0.02	0.98	97.54	95.94
		National Defense / Security	-0.02	0.98	100.00	98.39
Mistral-Small-24B-Instruct-2501	Economic	Market Economy	0.15	0.85	96.90	82.63
		Trade / Energy	0.01	0.99	91.20	90.47
		Labor	-0.14	0.86	96.88	83.25
		Welfare State	-0.02	0.98	97.58	96.01
	Sociocultural	Law and Order	-0.06	0.94	98.48	92.52
		Gender / Minorities / Equality	0.10	0.90	99.20	88.88
		International Relations	-0.02	0.98	97.54	95.94
		National Defense / Security	-0.04	0.96	100.00	95.97
Qwen3-0.6B-Base	Economic	Market Economy	0.38	0.62	93.80	58.17
		Trade / Energy	0.08	0.92	94.40	86.85
		Labor	-0.11	0.89	91.41	81.41
		Welfare State	0.05	0.95	92.74	88.25
	Sociocultural	Law and Order	0.06	0.94	94.70	88.96
		Gender / Minorities / Equality	0.15	0.85	89.60	75.98
		International Relations	0.03	0.97	83.61	80.87
		National Defense / Security	-0.18	0.82	91.13	74.96
Qwen3-0.6B	Economic	Market Economy	0.43	0.57	89.15	51.14
		Trade / Energy	0.15	0.85	93.60	79.37
		Labor	-0.06	0.94	84.38	79.10
		Welfare State	0.14	0.86	91.13	78.64
	Sociocultural	Law and Order	0.08	0.92	90.15	82.64
		Gender / Minorities / Equality	0.17	0.83	89.60	74.55
		International Relations	-0.10	0.90	78.69	70.95
		National Defense / Security	-0.18	0.82	87.10	71.64

Table 11 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
Qwen3-1.7B-Base	Economic	Market Economy	0.28	0.72	97.67	70.42
		Trade / Energy	0.07	0.93	95.20	88.35
		Labor	-0.13	0.88	94.53	82.71
		Welfare State	-0.03	0.97	98.39	95.21
	Sociocultural	Law and Order	0.15	0.85	94.70	80.35
		Gender / Minorities / Equality	0.14	0.86	96.00	82.94
		International Relations	-0.01	0.99	93.44	92.68
		National Defense / Security	-0.11	0.89	97.58	86.56
Qwen3-1.7B	Economic	Market Economy	0.26	0.74	93.80	69.80
		Trade / Energy	0.10	0.90	96.80	86.73
		Labor	-0.09	0.91	92.19	83.54
		Welfare State	0.08	0.92	96.77	88.97
	Sociocultural	Law and Order	0.06	0.94	90.91	85.40
		Gender / Minorities / Equality	0.12	0.88	90.40	79.55
		International Relations	-0.10	0.90	87.70	79.08
		National Defense / Security	-0.15	0.85	91.94	78.59
Qwen3-4B-Base	Economic	Market Economy	0.27	0.73	97.67	71.17
		Trade / Energy	0.09	0.91	95.20	86.82
		Labor	-0.19	0.81	96.09	78.08
		Welfare State	-0.13	0.87	98.39	85.69
	Sociocultural	Law and Order	0.05	0.95	96.21	91.84
		Gender / Minorities / Equality	0.09	0.91	95.20	86.82
		International Relations	-0.11	0.89	95.08	84.17
		National Defense / Security	0.02	0.98	95.97	94.42
Qwen3-4B	Economic	Market Economy	0.26	0.74	97.67	72.69
		Trade / Energy	0.09	0.91	93.60	85.36
		Labor	-0.19	0.81	91.41	74.27
		Welfare State	-0.05	0.95	93.55	89.02
	Sociocultural	Law and Order	0.08	0.92	93.18	86.12
		Gender / Minorities / Equality	0.11	0.89	90.40	80.28
		International Relations	-0.15	0.85	90.16	76.86
		National Defense / Security	-0.03	0.97	96.77	93.65
Qwen3-8B-Base	Economic	Market Economy	0.16	0.84	97.67	81.77
		Trade / Energy	0.06	0.94	96.80	90.60
		Labor	-0.21	0.79	98.44	77.67
		Welfare State	-0.08	0.92	99.19	91.19
	Sociocultural	Law and Order	0.08	0.92	98.48	91.02
		Gender / Minorities / Equality	0.05	0.95	94.40	89.87
		International Relations	-0.08	0.92	95.90	88.04
		National Defense / Security	-0.05	0.95	97.58	92.86
Qwen3-8B	Economic	Market Economy	0.19	0.81	96.90	78.12
		Trade / Energy	0.04	0.96	96.80	92.93
		Labor	-0.11	0.89	96.09	85.58
		Welfare State	0.01	0.99	97.58	96.79
	Sociocultural	Law and Order	0.10	0.90	98.48	88.79
		Gender / Minorities / Equality	0.10	0.90	92.80	83.15
		International Relations	-0.01	0.99	92.62	91.86
		National Defense / Security	0.05	0.95	96.77	92.09
Qwen3-14B-Base	Economic	Market Economy	0.21	0.79	97.67	77.23
		Trade / Energy	-0.01	0.99	97.60	96.82
		Labor	-0.20	0.80	98.44	78.44
		Welfare State	-0.02	0.98	100.00	98.39
	Sociocultural	Law and Order	0.00	1.00	98.48	98.48
		Gender / Minorities / Equality	0.00	1.00	97.60	97.60
		International Relations	0.13	0.87	96.72	84.04
		National Defense / Security	0.06	0.94	98.39	92.83

Table 11 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
Qwen3-14B	Economic	Market Economy	0.11	0.89	96.90	86.38
		Trade / Energy	-0.02	0.98	97.60	95.26
		Labor	-0.13	0.87	97.66	84.69
		Welfare State	0.08	0.92	99.19	91.19
	Sociocultural	Law and Order	0.02	0.98	98.48	96.25
		Gender / Minorities / Equality	0.04	0.96	96.00	92.16
		International Relations	0.08	0.92	95.08	87.29
		National Defense / Security	0.06	0.94	99.19	92.79
Qwen3-32B	Economic	Market Economy	0.07	0.93	97.67	90.86
		Trade / Energy	0.07	0.93	95.20	88.35
		Labor	-0.23	0.77	97.66	74.77
		Welfare State	0.08	0.92	99.19	91.19
	Sociocultural	Law and Order	-0.11	0.89	99.24	88.72
		Gender / Minorities / Equality	-0.12	0.88	98.40	86.59
		International Relations	0.13	0.87	96.72	84.04
		National Defense / Security	0.02	0.98	98.39	96.80
Midm-2.0-Mini-Instruct (2.3B)	Economic	Market Economy	0.02	0.98	99.22	96.92
		Trade / Energy	-0.13	0.87	96.00	83.71
		Labor	-0.14	0.86	100.00	85.94
		Welfare State	-0.06	0.94	99.19	93.59
	Sociocultural	Law and Order	-0.21	0.79	98.48	77.59
		Gender / Minorities / Equality	0.02	0.98	97.60	96.04
		International Relations	0.01	0.99	100.00	99.18
		National Defense / Security	0.05	0.95	100.00	95.16
Midm-2.0-Base-Instruct (11.5B)	Economic	Market Economy	0.04	0.96	100.00	96.12
		Trade / Energy	-0.21	0.79	96.00	76.03
		Labor	-0.13	0.88	100.00	87.50
		Welfare State	0.10	0.90	99.19	89.59
	Sociocultural	Law and Order	-0.19	0.81	99.24	80.45
		Gender / Minorities / Equality	0.03	0.97	97.60	94.48
		International Relations	0.05	0.95	100.00	95.08
		National Defense / Security	0.14	0.86	100.00	86.29
kanana-1.5-2.1b-base	Economic	Market Economy	0.05	0.95	100.00	94.57
		Trade / Energy	-0.12	0.88	96.80	85.18
		Labor	-0.14	0.86	99.22	85.27
		Welfare State	-0.05	0.95	100.00	95.16
	Sociocultural	Law and Order	-0.04	0.96	97.73	94.03
		Gender / Minorities / Equality	-0.04	0.96	96.00	92.16
		International Relations	-0.05	0.95	100.00	95.08
		National Defense / Security	0.00	1.00	100.00	100.00
kanana-1.5-2.1b-instruct-2505	Economic	Market Economy	-0.01	0.99	99.22	98.46
		Trade / Energy	-0.02	0.98	95.20	92.92
		Labor	-0.13	0.87	99.22	86.04
		Welfare State	0.02	0.98	100.00	98.39
	Sociocultural	Law and Order	-0.11	0.89	94.70	84.65
		Gender / Minorities / Equality	-0.04	0.96	96.80	92.93
		International Relations	0.02	0.98	99.18	97.55
		National Defense / Security	-0.02	0.98	100.00	98.39
kanana-1.5-8b-base	Economic	Market Economy	0.04	0.96	100.00	96.12
		Trade / Energy	-0.09	0.91	96.80	88.28
		Labor	-0.19	0.81	99.22	80.62
		Welfare State	-0.10	0.90	100.00	89.52
	Sociocultural	Law and Order	-0.02	0.98	97.73	96.25
		Gender / Minorities / Equality	-0.09	0.91	97.60	89.01
		International Relations	-0.05	0.95	100.00	95.08
		National Defense / Security	0.04	0.96	100.00	95.97

Table 11 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
kanana-1.5-8b-instruct-2505	Economic	Market Economy	-0.05	0.95	99.22	93.84
		Trade / Energy	-0.09	0.91	96.80	88.28
		Labor	-0.13	0.88	99.22	86.82
		Welfare State	-0.06	0.94	100.00	94.35
	Sociocultural	Law and Order	-0.06	0.94	98.48	92.52
		Gender / Minorities / Equality	-0.03	0.97	99.20	96.03
		International Relations	0.05	0.95	100.00	95.08
		National Defense / Security	0.06	0.94	99.19	92.79
EXAONE-4.0-1.2B	Economic	Market Economy	0.22	0.78	72.87	56.49
		Trade / Energy	0.10	0.90	83.20	74.55
		Labor	-0.06	0.94	82.03	76.90
		Welfare State	0.26	0.74	75.81	56.24
	Sociocultural	Law and Order	-0.15	0.85	87.88	74.56
		Gender / Minorities / Equality	0.26	0.74	75.20	55.35
		International Relations	-0.16	0.84	54.92	45.92
		National Defense / Security	-0.15	0.85	70.97	60.67
EXAONE-4.0-32B	Economic	Market Economy	0.17	0.83	94.57	78.44
		Trade / Energy	0.06	0.94	96.00	90.62
		Labor	-0.13	0.88	94.53	82.71
		Welfare State	0.08	0.92	95.16	87.49
	Sociocultural	Law and Order	-0.17	0.83	96.21	80.18
		Gender / Minorities / Equality	0.10	0.90	92.80	83.15
		International Relations	-0.03	0.97	91.80	88.79
		National Defense / Security	-0.03	0.97	95.16	92.09
HyperCLOVAX-SEED-Text-Instruct-0.5B	Economic	Market Economy	0.04	0.96	97.67	93.89
		Trade / Energy	0.01	0.99	88.00	87.30
		Labor	-0.02	0.98	96.88	95.36
		Welfare State	0.23	0.77	99.19	76.80
	Sociocultural	Law and Order	0.02	0.98	96.21	94.75
		Gender / Minorities / Equality	0.01	0.99	96.00	95.23
		International Relations	0.07	0.93	90.98	85.02
		National Defense / Security	0.11	0.89	97.58	86.56
HyperCLOVAX-SEED-Text-Instruct-1.5B	Economic	Market Economy	0.19	0.81	93.80	75.62
		Trade / Energy	0.11	0.89	90.40	80.28
		Labor	-0.06	0.94	84.38	79.10
		Welfare State	0.11	0.89	97.58	86.56
	Sociocultural	Law and Order	0.14	0.86	92.42	79.82
		Gender / Minorities / Equality	0.10	0.90	88.00	78.85
		International Relations	0.00	1.00	85.25	85.25
		National Defense / Security	0.03	0.97	94.35	91.31
A.X-4.0-Light (7B)	Economic	Market Economy	-0.05	0.95	100.00	95.35
		Trade / Energy	-0.23	0.77	99.20	76.19
		Labor	-0.14	0.86	99.22	85.27
		Welfare State	-0.02	0.98	98.39	96.01
	Sociocultural	Law and Order	-0.12	0.88	99.24	87.21
		Gender / Minorities / Equality	-0.12	0.88	100.00	88.00
		International Relations	0.03	0.97	100.00	96.72
		National Defense / Security	0.13	0.87	100.00	87.10
A.X-4.0 (72B)	Economic	Market Economy	-0.15	0.85	100.00	85.27
		Trade / Energy	-0.26	0.74	98.40	72.42
		Labor	-0.20	0.80	99.22	79.84
		Welfare State	-0.01	0.99	99.19	98.39
	Sociocultural	Law and Order	-0.06	0.94	100.00	93.94
		Gender / Minorities / Equality	-0.18	0.82	100.00	81.60
		International Relations	0.10	0.90	100.00	90.16
		National Defense / Security	0.05	0.95	100.00	95.16

Table 11 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
ko-gpt-trinity-1.2B-v0.5	Economic	Market Economy	-0.02	0.98	99.22	96.92
		Trade / Energy	-0.15	0.85	96.00	81.41
		Labor	-0.05	0.95	99.22	94.57
		Welfare State	0.00	1.00	100.00	100.00
	Sociocultural	Law and Order	-0.20	0.80	99.24	79.69
		Gender / Minorities / Equality	-0.18	0.82	98.40	80.29
		International Relations	0.00	1.00	99.18	99.18
		National Defense / Security	0.00	1.00	99.19	99.19
SOLAR-10.7B-v1.0	Economic	Market Economy	0.30	0.70	98.45	68.69
		Trade / Energy	-0.10	0.90	91.20	81.72
		Labor	-0.30	0.70	97.66	68.66
		Welfare State	0.15	0.85	98.39	84.11
	Sociocultural	Law and Order	-0.08	0.92	95.45	88.22
		Gender / Minorities / Equality	0.23	0.77	97.60	74.96
		International Relations	0.08	0.92	95.08	87.29
		National Defense / Security	0.03	0.97	99.19	95.99
SOLAR-10.7B-Instruct-v1.0	Economic	Market Economy	0.32	0.68	96.90	66.10
		Trade / Energy	-0.02	0.98	92.00	89.79
		Labor	-0.28	0.72	96.88	69.63
		Welfare State	0.14	0.86	98.39	84.90
	Sociocultural	Law and Order	0.05	0.95	94.70	90.39
		Gender / Minorities / Equality	0.22	0.78	97.60	76.52
		International Relations	0.00	1.00	95.08	95.08
		National Defense / Security	0.15	0.85	99.19	84.79

Table 12: Category-level evaluation results on the U.S. dataset translated into Korean.

Model	Axis	Category	Position	NS	LMS	ICAT
Llama-3.1-8B	Economic	Market Economy	0.00	1.00	95.42	95.42
		Trade / Energy	-0.21	0.79	96.27	76.15
		Labor	-0.22	0.78	99.20	77.77
		Welfare State	-0.20	0.80	93.08	74.46
	Sociocultural	Law and Order	-0.02	0.98	90.00	88.50
		Gender / Minorities / Equality	-0.27	0.73	90.00	65.45
		International Relations	0.23	0.77	99.25	76.29
		National Defense / Security	0.10	0.90	96.67	87.00
Llama-3.1-8B-Instruct	Economic	Market Economy	0.05	0.95	91.60	86.71
		Trade / Energy	-0.24	0.76	96.27	73.28
		Labor	-0.25	0.75	96.80	72.79
		Welfare State	-0.15	0.85	89.23	76.19
	Sociocultural	Law and Order	0.01	0.99	87.50	86.77
		Gender / Minorities / Equality	-0.22	0.78	91.82	71.79
		International Relations	0.16	0.84	99.25	82.96
		National Defense / Security	0.02	0.98	92.50	90.96
Llama-3.1-70B	Economic	Market Economy	-0.15	0.85	95.42	81.58
		Trade / Energy	-0.30	0.70	96.27	67.53
		Labor	-0.31	0.69	97.60	67.15
		Welfare State	-0.18	0.82	93.85	76.52
	Sociocultural	Law and Order	-0.08	0.92	89.17	81.74
		Gender / Minorities / Equality	-0.31	0.69	95.45	65.95
		International Relations	0.07	0.93	97.76	90.47
		National Defense / Security	0.03	0.97	95.83	92.64
Llama-3.1-70B-Instruct	Economic	Market Economy	-0.08	0.92	92.37	84.61
		Trade / Energy	-0.25	0.75	97.01	72.40
		Labor	-0.23	0.77	96.00	73.73
		Welfare State	-0.25	0.75	94.62	70.60
	Sociocultural	Law and Order	-0.07	0.93	90.00	84.00
		Gender / Minorities / Equality	-0.35	0.65	90.91	59.50
		International Relations	-0.04	0.96	97.01	92.67
		National Defense / Security	0.07	0.93	95.83	89.44
Llama-3.2-1B	Economic	Market Economy	0.11	0.89	88.55	79.09
		Trade / Energy	-0.18	0.82	98.51	80.86
		Labor	-0.09	0.91	92.00	83.90
		Welfare State	-0.03	0.97	86.92	84.25
	Sociocultural	Law and Order	0.15	0.85	85.00	72.25
		Gender / Minorities / Equality	-0.07	0.93	84.55	78.40
		International Relations	0.10	0.90	96.27	86.21
		National Defense / Security	0.03	0.97	90.83	87.81
Llama-3.2-1B-Instruct	Economic	Market Economy	0.18	0.82	79.39	65.45
		Trade / Energy	-0.11	0.89	96.27	85.49
		Labor	-0.02	0.98	92.80	90.57
		Welfare State	-0.06	0.94	88.46	83.02
	Sociocultural	Law and Order	0.10	0.90	83.33	75.00
		Gender / Minorities / Equality	-0.22	0.78	84.55	66.10
		International Relations	0.01	0.99	90.30	88.95
		National Defense / Security	0.18	0.82	85.00	69.42
Llama-3.2-3B	Economic	Market Economy	0.08	0.92	90.84	83.91
		Trade / Energy	-0.19	0.81	97.01	78.19
		Labor	-0.22	0.78	96.80	75.89
		Welfare State	-0.18	0.82	94.62	77.15
	Sociocultural	Law and Order	0.07	0.93	86.67	80.89
		Gender / Minorities / Equality	-0.25	0.75	91.82	68.45
		International Relations	0.16	0.84	98.51	83.07
		National Defense / Security	-0.02	0.98	93.33	91.78

Table 12 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
Llama-3.2-3B-Instruct	Economic	Market Economy	0.13	0.87	89.31	77.72
		Trade / Energy	-0.15	0.85	98.51	83.80
		Labor	-0.17	0.83	91.20	75.88
		Welfare State	-0.12	0.88	93.08	82.34
	Sociocultural	Law and Order	-0.03	0.97	82.50	79.75
		Gender / Minorities / Equality	-0.20	0.80	85.45	68.36
		International Relations	-0.07	0.93	96.27	89.08
		National Defense / Security	0.03	0.97	85.00	82.17
Mistral-7B-v0.3	Economic	Market Economy	0.18	0.82	77.86	64.19
		Trade / Energy	-0.18	0.82	96.27	79.03
		Labor	-0.23	0.77	97.60	74.96
		Welfare State	-0.15	0.85	89.23	75.50
	Sociocultural	Law and Order	-0.04	0.96	80.83	77.47
		Gender / Minorities / Equality	-0.08	0.92	81.82	75.12
		International Relations	0.10	0.90	91.04	81.53
		National Defense / Security	0.09	0.91	88.33	80.24
Mistral-7B-Instruct-v0.3	Economic	Market Economy	0.16	0.84	83.97	70.51
		Trade / Energy	-0.29	0.71	97.01	68.78
		Labor	-0.25	0.75	97.60	73.40
		Welfare State	-0.12	0.88	90.77	79.60
	Sociocultural	Law and Order	0.05	0.95	78.33	74.42
		Gender / Minorities / Equality	-0.13	0.87	85.45	74.58
		International Relations	0.09	0.91	91.04	82.89
		National Defense / Security	0.08	0.92	87.50	80.21
Mistral-Small-24B-Base-2501	Economic	Market Economy	0.05	0.95	93.13	88.15
		Trade / Energy	-0.37	0.63	98.51	61.75
		Labor	-0.34	0.66	96.00	62.98
		Welfare State	-0.16	0.84	96.15	80.62
	Sociocultural	Law and Order	-0.08	0.92	99.17	90.90
		Gender / Minorities / Equality	-0.33	0.67	93.64	62.99
		International Relations	-0.06	0.94	97.76	91.92
		National Defense / Security	-0.13	0.88	98.33	86.04
Mistral-Small-24B-Instruct-2501	Economic	Market Economy	0.00	1.00	93.13	93.13
		Trade / Energy	-0.37	0.63	97.76	61.28
		Labor	-0.33	0.67	94.40	63.44
		Welfare State	-0.15	0.85	96.15	81.36
	Sociocultural	Law and Order	-0.08	0.92	96.67	88.61
		Gender / Minorities / Equality	-0.21	0.79	90.91	71.90
		International Relations	-0.03	0.97	96.27	93.39
		National Defense / Security	-0.08	0.92	95.83	87.85
Qwen3-0.6B-Base	Economic	Market Economy	0.02	0.98	91.60	89.51
		Trade / Energy	-0.16	0.84	99.25	83.70
		Labor	-0.02	0.98	99.20	96.82
		Welfare State	-0.05	0.95	90.00	85.85
	Sociocultural	Law and Order	0.03	0.97	84.17	81.36
		Gender / Minorities / Equality	-0.13	0.87	80.00	69.82
		International Relations	0.02	0.98	96.27	94.11
		National Defense / Security	-0.08	0.92	80.83	74.10
Qwen3-0.6B	Economic	Market Economy	0.11	0.89	87.02	77.06
		Trade / Energy	-0.15	0.85	98.51	83.80
		Labor	0.12	0.88	96.00	84.48
		Welfare State	-0.06	0.94	94.62	88.79
	Sociocultural	Law and Order	-0.07	0.93	84.17	78.56
		Gender / Minorities / Equality	0.02	0.98	74.55	73.19
		International Relations	0.05	0.95	94.78	89.83
		National Defense / Security	0.07	0.93	81.67	76.22

Table 12 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
Qwen3-1.7B-Base	Economic	Market Economy	-0.02	0.98	93.13	91.00
		Trade / Energy	-0.24	0.76	98.51	74.98
		Labor	-0.20	0.80	98.40	78.72
		Welfare State	-0.03	0.97	91.54	88.72
	Sociocultural	Law and Order	-0.12	0.88	85.83	75.82
		Gender / Minorities / Equality	-0.11	0.89	85.45	76.13
		International Relations	0.04	0.96	98.51	94.83
		National Defense / Security	0.02	0.98	86.67	84.50
Qwen3-1.7B	Economic	Market Economy	0.01	0.99	87.79	87.12
		Trade / Energy	-0.28	0.72	97.01	69.50
		Labor	-0.09	0.91	96.00	87.55
		Welfare State	-0.12	0.88	89.23	78.25
	Sociocultural	Law and Order	-0.15	0.85	84.17	71.54
		Gender / Minorities / Equality	-0.04	0.96	80.00	77.09
		International Relations	0.07	0.93	96.27	89.08
		National Defense / Security	0.03	0.97	79.17	76.53
Qwen3-4B-Base	Economic	Market Economy	-0.01	0.99	96.95	96.21
		Trade / Energy	-0.25	0.75	98.51	73.51
		Labor	-0.23	0.77	98.40	75.57
		Welfare State	-0.14	0.86	89.23	76.88
	Sociocultural	Law and Order	-0.18	0.82	92.50	75.54
		Gender / Minorities / Equality	-0.20	0.80	91.82	73.45
		International Relations	0.06	0.94	98.51	92.63
		National Defense / Security	0.08	0.92	93.33	85.56
Qwen3-4B	Economic	Market Economy	-0.04	0.96	87.02	83.70
		Trade / Energy	-0.33	0.67	96.27	64.66
		Labor	-0.22	0.78	96.00	75.26
		Welfare State	-0.14	0.86	89.23	76.88
	Sociocultural	Law and Order	-0.08	0.92	85.83	78.68
		Gender / Minorities / Equality	-0.11	0.89	81.82	72.89
		International Relations	0.06	0.94	94.78	89.12
		National Defense / Security	0.10	0.90	88.33	79.50
Qwen3-8B-Base	Economic	Market Economy	0.04	0.96	97.71	93.98
		Trade / Energy	-0.18	0.82	97.76	80.25
		Labor	-0.18	0.82	99.20	81.74
		Welfare State	-0.12	0.88	89.23	78.93
	Sociocultural	Law and Order	-0.10	0.90	93.33	84.00
		Gender / Minorities / Equality	-0.26	0.74	92.73	68.28
		International Relations	0.09	0.91	98.51	89.69
		National Defense / Security	0.00	1.00	95.83	95.83
Qwen3-8B	Economic	Market Economy	-0.01	0.99	91.60	90.90
		Trade / Energy	-0.28	0.72	97.01	69.50
		Labor	-0.20	0.80	95.20	76.16
		Welfare State	-0.22	0.78	90.77	71.22
	Sociocultural	Law and Order	-0.05	0.95	89.17	84.71
		Gender / Minorities / Equality	-0.22	0.78	90.00	70.36
		International Relations	0.06	0.94	96.27	90.52
		National Defense / Security	0.04	0.96	91.67	87.85
Qwen3-14B-Base	Economic	Market Economy	-0.10	0.90	96.95	87.33
		Trade / Energy	-0.25	0.75	97.76	73.69
		Labor	-0.33	0.67	97.60	65.59
		Welfare State	-0.23	0.77	90.77	69.82
	Sociocultural	Law and Order	-0.17	0.83	94.17	78.47
		Gender / Minorities / Equality	-0.16	0.84	88.18	73.75
		International Relations	-0.05	0.95	99.25	94.07
		National Defense / Security	0.09	0.91	95.00	86.29

Table 12 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
Qwen3-14B	Economic	Market Economy	-0.11	0.89	95.42	84.49
		Trade / Energy	-0.28	0.72	97.01	70.23
		Labor	-0.31	0.69	96.00	66.05
		Welfare State	-0.26	0.74	90.77	67.03
	Sociocultural	Law and Order	-0.23	0.77	93.33	71.56
		Gender / Minorities / Equality	-0.07	0.93	87.27	80.93
		International Relations	-0.12	0.88	98.51	86.75
		National Defense / Security	0.10	0.90	92.50	83.25
Qwen3-32B	Economic	Market Economy	-0.16	0.84	92.37	77.56
		Trade / Energy	-0.33	0.67	97.76	65.66
		Labor	-0.19	0.81	96.00	77.57
		Welfare State	-0.37	0.63	90.00	56.77
	Sociocultural	Law and Order	-0.17	0.83	94.17	78.47
		Gender / Minorities / Equality	-0.22	0.78	90.00	70.36
		International Relations	-0.10	0.90	98.51	88.22
		National Defense / Security	0.05	0.95	92.50	87.88
Midm-2.0-Mini-Instruct (2.3B)	Economic	Market Economy	0.02	0.98	93.89	91.74
		Trade / Energy	-0.43	0.57	97.76	55.45
		Labor	-0.26	0.74	97.60	71.83
		Welfare State	-0.31	0.69	93.85	64.97
	Sociocultural	Law and Order	-0.13	0.88	95.00	83.13
		Gender / Minorities / Equality	-0.24	0.76	93.64	71.50
		International Relations	0.09	0.91	97.76	89.01
		National Defense / Security	0.13	0.87	99.17	85.94
Midm-2.0-Base-Instruct (11.5B)	Economic	Market Economy	-0.13	0.87	93.89	81.71
		Trade / Energy	-0.36	0.64	97.76	62.74
		Labor	-0.28	0.72	96.00	69.12
		Welfare State	-0.32	0.68	96.15	65.09
	Sociocultural	Law and Order	-0.07	0.93	96.67	90.22
		Gender / Minorities / Equality	-0.16	0.84	97.27	81.36
		International Relations	-0.03	0.97	97.76	94.84
		National Defense / Security	0.01	0.99	98.33	97.51
kanana-1.5-2.1b-base	Economic	Market Economy	-0.05	0.95	96.95	91.77
		Trade / Energy	-0.33	0.67	97.76	65.66
		Labor	-0.20	0.80	96.80	77.44
		Welfare State	-0.23	0.77	95.38	73.37
	Sociocultural	Law and Order	-0.15	0.85	93.33	79.33
		Gender / Minorities / Equality	-0.31	0.69	93.64	64.69
		International Relations	0.01	0.99	100.00	98.51
		National Defense / Security	0.00	1.00	96.67	96.67
kanana-1.5-2.1b-instruct-2505	Economic	Market Economy	-0.10	0.90	92.37	83.20
		Trade / Energy	-0.34	0.66	97.01	63.71
		Labor	-0.17	0.83	96.80	80.54
		Welfare State	-0.25	0.75	93.08	69.45
	Sociocultural	Law and Order	-0.16	0.84	90.83	76.45
		Gender / Minorities / Equality	-0.31	0.69	90.91	62.81
		International Relations	0.04	0.96	99.25	94.81
		National Defense / Security	0.08	0.93	95.83	88.65
kanana-1.5-8b-base	Economic	Market Economy	-0.11	0.89	96.18	85.17
		Trade / Energy	-0.33	0.67	97.76	65.66
		Labor	-0.23	0.77	97.60	74.96
		Welfare State	-0.35	0.65	95.38	61.63
	Sociocultural	Law and Order	-0.18	0.82	98.33	80.31
		Gender / Minorities / Equality	-0.32	0.68	92.73	63.22
		International Relations	0.03	0.97	99.25	96.29
		National Defense / Security	0.06	0.94	98.33	92.60

Table 12 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
kanana-1.5-8b-instruct-2505	Economic	Market Economy	-0.16	0.84	93.13	78.20
		Trade / Energy	-0.34	0.66	97.76	64.20
		Labor	-0.26	0.74	97.60	71.83
		Welfare State	-0.29	0.71	96.15	68.05
	Sociocultural	Law and Order	-0.13	0.87	95.00	82.33
		Gender / Minorities / Equality	-0.33	0.67	91.82	61.77
		International Relations	0.12	0.88	99.25	87.40
		National Defense / Security	0.10	0.90	96.67	87.00
EXAONE-4.0-1.2B	Economic	Market Economy	0.01	0.99	65.65	65.15
		Trade / Energy	-0.04	0.96	85.82	82.62
		Labor	-0.01	0.99	79.20	78.57
		Welfare State	0.03	0.97	72.31	70.08
	Sociocultural	Law and Order	0.00	1.00	70.00	70.00
		Gender / Minorities / Equality	-0.04	0.96	63.64	61.32
		International Relations	0.03	0.97	78.36	76.02
		National Defense / Security	0.13	0.87	63.33	54.89
EXAONE-4.0-32B	Economic	Market Economy	-0.01	0.99	89.31	88.63
		Trade / Energy	-0.22	0.78	97.76	75.87
		Labor	-0.20	0.80	92.00	73.60
		Welfare State	-0.26	0.74	90.00	66.46
	Sociocultural	Law and Order	-0.10	0.90	95.83	86.25
		Gender / Minorities / Equality	-0.25	0.75	79.09	58.96
		International Relations	0.01	0.99	95.52	94.10
		National Defense / Security	0.08	0.92	88.33	80.97
HyperCLOVAX-SEED-Text-Instruct-0.5B	Economic	Market Economy	-0.08	0.92	89.31	82.50
		Trade / Energy	-0.06	0.94	94.78	89.12
		Labor	-0.04	0.96	87.20	83.71
		Welfare State	-0.12	0.88	83.85	73.53
	Sociocultural	Law and Order	0.05	0.95	91.67	87.08
		Gender / Minorities / Equality	-0.22	0.78	88.18	68.94
		International Relations	0.19	0.81	91.79	74.67
		National Defense / Security	0.08	0.93	87.50	80.94
HyperCLOVAX-SEED-Text-Instruct-1.5B	Economic	Market Economy	-0.11	0.89	83.97	74.35
		Trade / Energy	-0.13	0.87	94.03	81.40
		Labor	-0.06	0.94	85.60	80.81
		Welfare State	-0.11	0.89	83.08	74.13
	Sociocultural	Law and Order	-0.05	0.95	83.33	79.17
		Gender / Minorities / Equality	-0.18	0.82	79.09	64.71
		International Relations	0.13	0.87	88.81	77.54
		National Defense / Security	-0.03	0.97	84.17	81.36
A.X-4.0-Light (7B)	Economic	Market Economy	-0.15	0.85	93.13	79.62
		Trade / Energy	-0.27	0.73	100.00	73.13
		Labor	-0.26	0.74	96.00	70.66
		Welfare State	-0.20	0.80	93.08	74.46
	Sociocultural	Law and Order	-0.07	0.93	98.33	91.78
		Gender / Minorities / Equality	-0.24	0.76	92.73	70.81
		International Relations	-0.06	0.94	97.76	91.92
		National Defense / Security	0.00	1.00	99.17	99.17
A.X-4.0 (72B)	Economic	Market Economy	-0.28	0.72	96.18	69.02
		Trade / Energy	-0.34	0.66	99.25	65.18
		Labor	-0.31	0.69	99.20	68.25
		Welfare State	-0.37	0.63	96.92	61.14
	Sociocultural	Law and Order	-0.10	0.90	98.33	88.50
		Gender / Minorities / Equality	-0.33	0.67	96.36	64.83
		International Relations	-0.15	0.85	100.00	85.07
		National Defense / Security	0.05	0.95	99.17	94.21

Table 12 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
ko-gpt-trinity-1.2B-v0.5	Economic	Market Economy	-0.10	0.90	93.13	83.89
		Trade / Energy	-0.25	0.75	99.25	74.07
		Labor	-0.15	0.85	92.80	78.69
		Welfare State	-0.38	0.62	96.92	59.64
	Sociocultural	Law and Order	-0.08	0.92	87.50	80.21
		Gender / Minorities / Equality	-0.27	0.73	90.00	65.45
		International Relations	0.00	1.00	97.76	97.76
		National Defense / Security	0.08	0.92	95.83	87.85
SOLAR-10.7B-v1.0	Economic	Market Economy	-0.07	0.93	87.79	81.76
		Trade / Energy	-0.21	0.79	99.25	78.51
		Labor	-0.15	0.85	97.60	82.76
		Welfare State	-0.02	0.98	96.15	94.67
	Sociocultural	Law and Order	-0.05	0.95	91.67	87.08
		Gender / Minorities / Equality	-0.31	0.69	90.00	62.18
		International Relations	0.27	0.73	96.27	70.41
		National Defense / Security	-0.02	0.98	95.83	94.24
SOLAR-10.7B-Instruct-v1.0	Economic	Market Economy	-0.07	0.93	87.02	81.04
		Trade / Energy	-0.30	0.70	97.76	68.58
		Labor	-0.20	0.80	96.80	77.44
		Welfare State	-0.14	0.86	95.38	82.18
	Sociocultural	Law and Order	-0.03	0.97	91.67	88.61
		Gender / Minorities / Equality	-0.37	0.63	86.36	54.17
		International Relations	0.12	0.88	98.51	86.75
		National Defense / Security	-0.08	0.92	94.17	86.32

Table 13: Category-level evaluation results on the South Korean dataset translated into English.

Model	Axis	Category	Position	NS	LMS	ICAT
Llama-3.1-8B	Economic	Market Economy	-0.22	0.78	96.90	75.12
		Trade / Energy	-0.35	0.65	92.80	60.13
		Labor	-0.19	0.81	97.66	79.35
		Welfare State	-0.21	0.79	95.97	75.85
	Sociocultural	Law and Order	-0.11	0.89	94.70	84.65
		Gender / Minorities / Equality	-0.04	0.96	95.20	91.39
		International Relations	-0.05	0.95	99.18	94.30
		National Defense / Security	-0.05	0.95	95.97	91.32
Llama-3.1-8B-Instruct	Economic	Market Economy	-0.22	0.78	96.90	75.12
		Trade / Energy	-0.30	0.70	92.80	65.33
		Labor	-0.20	0.80	96.88	77.20
		Welfare State	-0.10	0.90	94.35	85.22
	Sociocultural	Law and Order	-0.11	0.89	93.18	83.30
		Gender / Minorities / Equality	0.01	0.99	96.00	95.23
		International Relations	-0.07	0.93	97.54	91.14
		National Defense / Security	-0.05	0.95	95.16	90.56
Llama-3.1-70B	Economic	Market Economy	-0.22	0.78	96.12	74.51
		Trade / Energy	-0.39	0.61	93.60	56.91
		Labor	-0.27	0.73	97.66	71.72
		Welfare State	-0.24	0.76	98.39	74.58
	Sociocultural	Law and Order	-0.09	0.91	97.73	88.84
		Gender / Minorities / Equality	-0.01	0.99	95.20	94.44
		International Relations	-0.10	0.90	99.18	89.42
		National Defense / Security	-0.02	0.98	99.19	97.59
Llama-3.1-70B-Instruct	Economic	Market Economy	-0.22	0.78	96.90	75.12
		Trade / Energy	-0.34	0.66	91.20	59.83
		Labor	-0.23	0.77	97.66	74.77
		Welfare State	-0.21	0.79	96.77	76.48
	Sociocultural	Law and Order	-0.12	0.88	97.73	85.88
		Gender / Minorities / Equality	0.02	0.98	96.80	94.48
		International Relations	-0.10	0.90	98.36	88.69
		National Defense / Security	-0.05	0.95	97.58	92.86
Llama-3.2-1B	Economic	Market Economy	-0.07	0.93	95.35	88.70
		Trade / Energy	-0.26	0.74	93.60	68.89
		Labor	-0.23	0.77	96.09	73.57
		Welfare State	-0.06	0.94	94.35	88.27
	Sociocultural	Law and Order	-0.08	0.92	93.18	86.12
		Gender / Minorities / Equality	0.04	0.96	96.80	92.93
		International Relations	-0.19	0.81	95.90	77.82
		National Defense / Security	-0.17	0.83	92.74	77.04
Llama-3.2-1B-Instruct	Economic	Market Economy	-0.05	0.95	90.70	85.78
		Trade / Energy	-0.12	0.88	88.00	77.44
		Labor	-0.14	0.86	92.97	79.90
		Welfare State	-0.13	0.87	87.10	75.86
	Sociocultural	Law and Order	-0.11	0.89	89.39	79.91
		Gender / Minorities / Equality	0.04	0.96	92.80	89.09
		International Relations	-0.16	0.84	93.44	78.12
		National Defense / Security	-0.23	0.77	91.13	70.55
Llama-3.2-3B	Economic	Market Economy	-0.09	0.91	97.67	88.59
		Trade / Energy	-0.27	0.73	92.00	66.98
		Labor	-0.20	0.80	96.09	76.57
		Welfare State	-0.13	0.87	93.55	81.48
	Sociocultural	Law and Order	-0.05	0.95	90.91	86.78
		Gender / Minorities / Equality	0.04	0.96	96.80	92.93
		International Relations	-0.08	0.92	95.90	88.04
		National Defense / Security	-0.11	0.89	93.55	82.99

Table 13 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
Llama-3.2-3B-Instruct	Economic	Market Economy	-0.13	0.87	93.80	81.44
		Trade / Energy	-0.17	0.83	88.00	73.22
		Labor	-0.22	0.78	92.19	72.02
		Welfare State	-0.11	0.89	90.32	80.12
	Sociocultural	Law and Order	-0.09	0.91	90.91	82.64
		Gender / Minorities / Equality	0.02	0.98	94.40	92.13
		International Relations	0.03	0.97	93.44	90.38
		National Defense / Security	-0.21	0.79	91.13	72.02
Mistral-7B-v0.3	Economic	Market Economy	-0.07	0.93	99.22	92.30
		Trade / Energy	-0.21	0.79	93.60	74.13
		Labor	-0.11	0.89	96.88	86.28
		Welfare State	-0.19	0.81	95.97	78.17
	Sociocultural	Law and Order	0.03	0.97	96.21	93.30
		Gender / Minorities / Equality	0.06	0.94	97.60	92.13
		International Relations	-0.05	0.95	100.00	95.08
		National Defense / Security	-0.10	0.90	96.77	87.41
Mistral-7B-Instruct-v0.3	Economic	Market Economy	-0.09	0.91	99.22	90.76
		Trade / Energy	-0.20	0.80	94.40	75.52
		Labor	-0.09	0.91	96.88	87.79
		Welfare State	-0.21	0.79	95.97	75.85
	Sociocultural	Law and Order	-0.06	0.94	96.97	91.09
		Gender / Minorities / Equality	0.01	0.99	97.60	96.82
		International Relations	-0.06	0.94	100.00	94.26
		National Defense / Security	-0.11	0.89	97.58	86.56
Mistral-Small-24B-Base-2501	Economic	Market Economy	-0.16	0.84	97.67	81.77
		Trade / Energy	-0.33	0.67	91.20	61.29
		Labor	-0.25	0.75	97.66	73.24
		Welfare State	-0.19	0.81	96.77	78.04
	Sociocultural	Law and Order	-0.08	0.92	96.97	89.62
		Gender / Minorities / Equality	-0.01	0.99	96.00	95.23
		International Relations	-0.02	0.98	99.18	97.55
		National Defense / Security	0.05	0.95	95.16	90.56
Mistral-Small-24B-Instruct-2501	Economic	Market Economy	-0.19	0.81	96.90	78.87
		Trade / Energy	-0.31	0.69	91.20	62.75
		Labor	-0.19	0.81	97.66	79.35
		Welfare State	-0.16	0.84	95.16	79.81
	Sociocultural	Law and Order	-0.08	0.92	96.21	88.19
		Gender / Minorities / Equality	-0.02	0.98	95.20	92.92
		International Relations	-0.05	0.95	99.18	94.30
		National Defense / Security	0.05	0.95	95.97	91.32
Qwen3-0.6B-Base	Economic	Market Economy	-0.21	0.79	93.80	74.17
		Trade / Energy	-0.31	0.69	89.60	61.64
		Labor	-0.22	0.78	94.53	73.85
		Welfare State	-0.18	0.82	92.74	76.29
	Sociocultural	Law and Order	0.02	0.98	92.42	91.02
		Gender / Minorities / Equality	0.06	0.94	95.20	89.87
		International Relations	-0.03	0.97	95.90	92.76
		National Defense / Security	-0.05	0.95	90.32	85.95
Qwen3-0.6B	Economic	Market Economy	-0.19	0.81	91.47	73.75
		Trade / Energy	-0.33	0.67	87.20	58.60
		Labor	-0.25	0.75	90.63	67.97
		Welfare State	-0.23	0.77	91.13	70.55
	Sociocultural	Law and Order	-0.03	0.97	91.67	88.89
		Gender / Minorities / Equality	-0.02	0.98	92.00	90.53
		International Relations	-0.02	0.98	93.44	91.91
		National Defense / Security	-0.16	0.84	88.71	74.40

Table 13 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
Qwen3-1.7B-Base	Economic	Market Economy	-0.16	0.84	95.35	80.57
		Trade / Energy	-0.24	0.76	90.40	68.70
		Labor	-0.19	0.81	95.31	77.44
		Welfare State	-0.27	0.73	92.74	67.31
	Sociocultural	Law and Order	-0.08	0.92	92.42	85.42
		Gender / Minorities / Equality	0.01	0.99	95.20	94.44
		International Relations	-0.11	0.89	96.72	85.62
		National Defense / Security	-0.04	0.96	94.35	90.55
Qwen3-1.7B	Economic	Market Economy	-0.18	0.82	91.47	75.16
		Trade / Energy	-0.26	0.74	86.40	63.59
		Labor	-0.17	0.83	92.97	76.99
		Welfare State	-0.25	0.75	89.52	67.14
	Sociocultural	Law and Order	0.02	0.98	90.15	88.10
		Gender / Minorities / Equality	-0.01	0.99	90.40	89.68
		International Relations	-0.08	0.92	95.08	87.29
		National Defense / Security	-0.06	0.94	90.32	84.50
Qwen3-4B-Base	Economic	Market Economy	-0.21	0.79	95.35	75.39
		Trade / Energy	-0.38	0.62	92.00	57.41
		Labor	-0.19	0.81	96.88	78.71
		Welfare State	-0.24	0.76	91.94	69.69
	Sociocultural	Law and Order	-0.09	0.91	94.70	86.09
		Gender / Minorities / Equality	-0.01	0.99	96.80	96.03
		International Relations	-0.07	0.93	96.72	90.38
		National Defense / Security	-0.10	0.90	93.55	84.50
Qwen3-4B	Economic	Market Economy	-0.22	0.78	93.80	72.71
		Trade / Energy	-0.28	0.72	89.60	64.51
		Labor	-0.20	0.80	93.75	74.71
		Welfare State	-0.24	0.76	87.90	66.64
	Sociocultural	Law and Order	0.05	0.95	90.15	86.05
		Gender / Minorities / Equality	-0.01	0.99	92.00	91.26
		International Relations	-0.03	0.97	95.08	91.96
		National Defense / Security	-0.19	0.81	92.74	74.79
Qwen3-8B-Base	Economic	Market Economy	-0.27	0.73	94.57	68.91
		Trade / Energy	-0.28	0.72	91.20	65.66
		Labor	-0.23	0.77	97.66	74.77
		Welfare State	-0.24	0.76	91.94	69.69
	Sociocultural	Law and Order	0.01	0.99	94.70	93.98
		Gender / Minorities / Equality	0.03	0.97	96.00	92.93
		International Relations	-0.02	0.98	98.36	96.75
		National Defense / Security	-0.11	0.89	95.97	85.13
Qwen3-8B	Economic	Market Economy	-0.22	0.78	94.57	73.31
		Trade / Energy	-0.33	0.67	89.60	60.21
		Labor	-0.20	0.80	96.09	76.57
		Welfare State	-0.23	0.77	92.74	71.80
	Sociocultural	Law and Order	-0.05	0.95	92.42	88.22
		Gender / Minorities / Equality	0.02	0.98	94.40	92.13
		International Relations	-0.04	0.96	98.36	94.33
		National Defense / Security	-0.19	0.81	93.55	76.20
Qwen3-14B-Base	Economic	Market Economy	-0.26	0.74	96.90	72.11
		Trade / Energy	-0.34	0.66	91.20	60.56
		Labor	-0.18	0.82	97.66	80.11
		Welfare State	-0.29	0.71	91.94	65.24
	Sociocultural	Law and Order	-0.08	0.92	95.45	88.22
		Gender / Minorities / Equality	-0.02	0.98	97.60	95.26
		International Relations	-0.07	0.93	97.54	90.35
		National Defense / Security	-0.13	0.87	95.16	82.88

Table 13 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
Qwen3-14B	Economic	Market Economy	-0.20	0.80	96.12	76.75
		Trade / Energy	-0.41	0.59	89.60	53.04
		Labor	-0.23	0.77	95.31	72.97
		Welfare State	-0.16	0.84	93.55	78.46
	Sociocultural	Law and Order	-0.13	0.87	96.21	83.82
		Gender / Minorities / Equality	0.06	0.94	96.00	89.86
		International Relations	-0.05	0.95	97.54	92.74
		National Defense / Security	-0.10	0.90	94.35	85.22
Qwen3-32B	Economic	Market Economy	-0.18	0.82	96.12	78.99
		Trade / Energy	-0.31	0.69	92.00	63.30
		Labor	-0.25	0.75	96.09	72.07
		Welfare State	-0.18	0.82	93.55	76.95
	Sociocultural	Law and Order	-0.08	0.92	96.21	88.92
		Gender / Minorities / Equality	0.04	0.96	96.00	92.16
		International Relations	-0.03	0.97	99.18	95.93
		National Defense / Security	-0.03	0.97	95.97	92.87
Midm-2.0-Mini-Instruct (2.3B)	Economic	Market Economy	-0.15	0.85	96.90	82.63
		Trade / Energy	-0.30	0.70	92.00	64.77
		Labor	-0.28	0.72	93.75	67.38
		Welfare State	-0.19	0.81	93.55	75.44
	Sociocultural	Law and Order	-0.11	0.89	93.18	83.30
		Gender / Minorities / Equality	0.00	1.00	93.60	93.60
		International Relations	-0.02	0.98	98.36	96.75
		National Defense / Security	-0.21	0.79	95.97	75.85
Midm-2.0-Base-Instruct (11.5B)	Economic	Market Economy	-0.10	0.90	98.45	88.53
		Trade / Energy	-0.26	0.74	93.60	68.89
		Labor	-0.23	0.77	99.22	76.74
		Welfare State	-0.23	0.77	95.97	73.52
	Sociocultural	Law and Order	-0.04	0.96	95.45	91.84
		Gender / Minorities / Equality	-0.07	0.93	95.20	88.35
		International Relations	-0.02	0.98	98.36	96.75
		National Defense / Security	-0.13	0.87	96.77	84.29
kanana-1.5-2.1b-base	Economic	Market Economy	-0.15	0.85	96.12	81.97
		Trade / Energy	-0.30	0.70	90.40	63.64
		Labor	-0.23	0.77	95.31	72.97
		Welfare State	-0.17	0.83	92.74	77.04
	Sociocultural	Law and Order	-0.09	0.91	90.91	82.64
		Gender / Minorities / Equality	-0.04	0.96	97.60	93.70
		International Relations	-0.13	0.87	95.90	83.32
		National Defense / Security	-0.10	0.90	92.74	83.77
kanana-1.5-2.1b-instruct-2505	Economic	Market Economy	-0.07	0.93	97.67	90.86
		Trade / Energy	-0.30	0.70	92.80	65.33
		Labor	-0.23	0.77	94.53	72.38
		Welfare State	-0.13	0.87	92.74	80.78
	Sociocultural	Law and Order	-0.02	0.98	94.70	93.26
		Gender / Minorities / Equality	-0.01	0.99	96.80	96.03
		International Relations	-0.08	0.92	98.36	90.30
		National Defense / Security	-0.10	0.90	93.55	84.50
kanana-1.5-8b-base	Economic	Market Economy	-0.15	0.85	97.67	83.29
		Trade / Energy	-0.31	0.69	91.20	62.75
		Labor	-0.20	0.80	97.66	77.82
		Welfare State	-0.16	0.84	94.35	79.14
	Sociocultural	Law and Order	-0.11	0.89	93.18	83.30
		Gender / Minorities / Equality	0.01	0.99	94.40	93.64
		International Relations	-0.02	0.98	96.72	95.14
		National Defense / Security	-0.08	0.92	91.94	84.52

Table 13 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
kanana-1.5-8b-instruct-2505	Economic	Market Economy	-0.05	0.95	97.67	92.37
		Trade / Energy	-0.31	0.69	84.80	58.34
		Labor	-0.17	0.83	95.31	78.93
		Welfare State	-0.19	0.81	93.55	76.20
	Sociocultural	Law and Order	-0.12	0.88	90.91	79.89
		Gender / Minorities / Equality	0.04	0.96	92.80	89.09
		International Relations	0.07	0.93	98.36	91.91
		National Defense / Security	-0.03	0.97	92.74	89.75
EXAONE-4.0-1.2B	Economic	Market Economy	-0.24	0.76	62.79	47.70
		Trade / Energy	-0.22	0.78	56.00	43.90
		Labor	-0.27	0.73	64.84	47.62
		Welfare State	-0.27	0.73	68.55	49.75
	Sociocultural	Law and Order	-0.02	0.98	74.24	73.12
		Gender / Minorities / Equality	-0.03	0.97	76.80	74.34
		International Relations	0.00	1.00	69.67	69.67
		National Defense / Security	-0.10	0.90	68.55	61.91
EXAONE-4.0-32B	Economic	Market Economy	-0.27	0.73	96.90	70.61
		Trade / Energy	-0.30	0.70	88.80	62.52
		Labor	-0.25	0.75	94.53	70.90
		Welfare State	-0.17	0.83	95.16	79.05
	Sociocultural	Law and Order	-0.14	0.86	94.70	81.78
		Gender / Minorities / Equality	0.02	0.98	94.40	92.13
		International Relations	-0.05	0.95	98.36	93.52
		National Defense / Security	-0.11	0.89	95.16	84.42
HyperCLOVAX-SEED-Text-Instruct-0.5B	Economic	Market Economy	-0.12	0.88	90.70	80.15
		Trade / Energy	-0.20	0.80	90.40	72.32
		Labor	-0.24	0.76	92.97	70.45
		Welfare State	-0.06	0.94	91.13	85.98
	Sociocultural	Law and Order	-0.02	0.98	89.39	88.04
		Gender / Minorities / Equality	-0.02	0.98	94.40	92.89
		International Relations	-0.07	0.93	95.90	89.61
		National Defense / Security	-0.26	0.74	91.13	67.61
HyperCLOVAX-SEED-Text-Instruct-1.5B	Economic	Market Economy	-0.18	0.82	86.05	70.70
		Trade / Energy	-0.27	0.73	88.00	64.06
		Labor	-0.20	0.80	93.75	75.44
		Welfare State	-0.14	0.86	89.52	77.24
	Sociocultural	Law and Order	-0.09	0.91	87.12	79.20
		Gender / Minorities / Equality	-0.02	0.98	95.20	92.92
		International Relations	-0.10	0.90	94.26	84.99
		National Defense / Security	-0.19	0.81	91.13	73.49
A.X-4.0-Light (7B)	Economic	Market Economy	-0.12	0.88	97.67	85.56
		Trade / Energy	-0.30	0.70	86.40	60.83
		Labor	-0.20	0.80	92.97	74.08
		Welfare State	-0.13	0.87	91.13	79.37
	Sociocultural	Law and Order	0.08	0.92	95.45	88.22
		Gender / Minorities / Equality	-0.01	0.99	97.60	96.82
		International Relations	-0.05	0.95	96.72	91.96
		National Defense / Security	-0.08	0.92	95.97	88.23
A.X-4.0 (72B)	Economic	Market Economy	-0.27	0.73	98.45	71.74
		Trade / Energy	-0.34	0.66	93.60	61.40
		Labor	-0.20	0.80	98.44	78.44
		Welfare State	-0.18	0.82	98.39	80.93
	Sociocultural	Law and Order	-0.02	0.98	99.24	97.74
		Gender / Minorities / Equality	-0.18	0.82	99.20	80.95
		International Relations	-0.04	0.96	100.00	95.90
		National Defense / Security	-0.03	0.97	99.19	95.99

Table 13 continued from previous page

Model	Axis	Category	Position	NS	LMS	ICAT
ko-gpt-trinity-1.2B-v0.5	Economic	Market Economy	-0.01	0.99	99.22	98.46
		Trade / Energy	0.15	0.85	100.00	84.80
		Labor	-0.05	0.95	98.44	93.82
		Welfare State	-0.19	0.81	99.19	79.99
	Sociocultural	Law and Order	0.11	0.89	99.24	88.72
		Gender / Minorities / Equality	-0.18	0.82	98.40	80.29
		International Relations	-0.11	0.89	100.00	88.52
		National Defense / Security	-0.29	0.71	100.00	70.97
SOLAR-10.7B-v1.0	Economic	Market Economy	-0.12	0.88	99.22	87.69
		Trade / Energy	-0.20	0.80	92.80	74.24
		Labor	-0.14	0.86	98.44	84.59
		Welfare State	-0.18	0.82	97.58	80.27
	Sociocultural	Law and Order	-0.02	0.98	97.73	96.25
		Gender / Minorities / Equality	-0.02	0.98	98.40	96.04
		International Relations	-0.10	0.90	100.00	90.16
		National Defense / Security	-0.08	0.92	96.77	88.97
SOLAR-10.7B-Instruct-v1.0	Economic	Market Economy	0.04	0.96	99.22	95.38
		Trade / Energy	-0.23	0.77	92.80	71.27
		Labor	-0.09	0.91	97.66	88.50
		Welfare State	-0.18	0.82	96.77	79.60
	Sociocultural	Law and Order	0.03	0.97	96.21	93.30
		Gender / Minorities / Equality	0.01	0.99	96.80	96.03
		International Relations	-0.02	0.98	100.00	98.36
		National Defense / Security	-0.05	0.95	95.16	90.56