HebID: Detecting Social Identities in Hebrew-language Political Text

Guy Mor-Lan[1], Naama Rivlin-Angert[1], Yael R. Kaplan[2], Tamir Sheafer[1], Shaul R. Shenhav[1]
[1] The Hebrew University of Jerusalem, [2] The Open University of Israel
guy.mor@mail.huji.ac.il

Abstract

Political language is deeply intertwined with social identities. While social identities are often shaped by specific cultural contexts and expressed through particular uses of language, existing datasets for group and identity detection are predominantly English-centric, single-label and focus on coarse identity categories. We introduce HebID, the first multilabel Hebrew corpus for social identity detection: 5,536 sentences from Israeli politicians' Facebook posts (Dec 2018–Apr 2021), manually annotated for twelve nuanced social identities (e.g., Rightist, Ultra-Orthodox, Socially-oriented) grounded by survey data. We benchmark multilabel and single-label encoders alongside 2B–9B-parameter seq2seq LLMs, finding that Hebrew-tuned LLMs provide the best results (macro-F_1 = 0.74). We apply our classifier to politicians' Facebook posts and parliamentary speeches, evaluating differences in popularity, temporal trends, clustering patterns, and gender-related variations in identity expression. We utilize identity choices from a national public survey, enabling a comparison between identities portrayed in elite discourse and the public's identity priorities. HebID provides a comprehensive foundation for studying social identities in Hebrew and can serve as a model for similar research in other non-English political contexts.[1]

1 Introduction

Social identities—such as political ideology, religious affiliation, or demographic group membership—are powerful drivers of political behavior and public discourse (Tajfel and Turner, 1979a,b). Yet existing NLP resources for identity detection remain almost entirely English-focused, single-label, and rely on coarse categories (e.g., party or ethnicity). In this work, we introduce **HebID**, the first publicly released Hebrew dataset for fine-grained, multilabel social identity detection in political text, grounded in both domain expertise and large-scale survey evidence. We utilized 12 waves of a national survey (N = 1,769) each of which included questions designed to identify the twelve expert-defined social identities most salient to Israeli citizens (e.g., Rightist, Ultra-Orthodox, Socially-oriented). These survey-derived categories were then used to guide the annotation of 5,536 sentences sampled from Facebook posts by Israeli politicians (December 2018–April 2021), ensuring that our labels reflect real-world identity salience.

We benchmark a suite of modern architectures on the Facebook data—(i) multilabel encoder models; (ii) single-label encoder models; and (iii) 2B–9B-parameter seq2seq LLMs, finding that Hebrew-tuned seq2seq models (specifically, DictaLM2.0) achieve the highest macro-F_1 (0.743), outperforming encoder-only baselines by over six points. We then assess cross-genre generalization by applying our best model to 500 Knesset speech excerpts, achieving a comparable macro-F_1 of 0.72.

Finally, we link three complementary sources in our analysis of how identities behave: politicians' Facebook posts, parliamentary speeches from the Knesset floor, and survey responses that showcase public identification. We (1) compare identity prevalence and correlations between social media, parliamentary speech, and the public; (2) document how identity discourse surges around election cycles; (3) uncover coherent "bundles" of co-occurring identities; and (4) quantify gender-related variation in identity expression.

Our contributions are:

- **Dataset**: a multilabel Hebrew corpus of 5,536 sentences annotated for 12 social identities, empirically grounded in survey data.
- **Benchmarks**: comprehensive comparison of encoder-only and seq2seq LLMs, showing the superiority of the latter.

[1] https://github.com/guymorlan/hebid/

- **Cross-genre evaluation**: demonstration of model generalization to parliamentary speech.
- **Sociolinguistic analysis**: novel insights into identity popularity, temporal dynamics, bundling, and gendered expression across multiple sources.

2 Related Work

Automatic analysis of social identity language in political text intersects several NLP subfields: group reference detection, framing and stance classification, and ideological position inference. However, existing resources remain limited in granularity, language coverage, and multilabel capacity, leaving a clear gap that our Hebrew dataset fills.

Group reference detection. Early work extended named-entity recognition to capture social groups as entities. Zanotto et al. (2024) introduce GRIT, annotating Italian news and parliamentary text for spans referring to demographic, national, or partisan groups, and fine-tune BERT to identify them. Licht and Sczepanski (2024) apply a similar span-labeling approach to British parliamentary debates, quantifying how often politicians mention particular social groups. These studies demonstrate the feasibility of automatic group mention detection, but remain monolingual, single-label, and limited to explicit group mentions.

Framing and stance. Beyond mentions, understanding how groups are portrayed is crucial. The *Us vs. Them* corpus (Huguet Cabot et al., 2021) annotates Reddit comments for target group, stance (supportive vs. hostile), and emotion, using a multi-task RoBERTa model. Card et al. (2022) study 140 years of U.S. congressional speeches on immigration, combining sentiment classification with custom frame lexica to reveal evolving and polarized frames. The Media Frames Corpus (Card et al., 2015) provides frame annotations (including cultural identity frames) for news articles. These resources focus on single-label stance or framing, and predominantly English texts.

Ideological position inference. Separate but related is the task of inferring latent political ideology. Thomas et al. (2006) use SVMs on floor debate transcripts to classify support/opposition. Iyyer et al. (2014) develop neural networks to predict left/right alignment of speeches. More recent prompt-based methods with LLMs (Mishra et al., 2023) achieve zero-shot ideological scoring. While these approaches infer broad ideological leanings, they do not detect explicit identity mentions or support multilabel identity categories.

In-group/out-group rhetoric. Discourse-level signals such as pronoun clusivity and coded language reveal populist identity appeals. Rehbein and Ruppenhofer (2022) annotate inclusive vs. exclusive uses of “we” in German parliamentary debates and train classifiers to disambiguate referents, highlighting *us-versus-them* framing. Mendelsohn et al. (2023) curate a dogwhistle lexicon and show that pretrained models struggle to detect covert slurs without knowledge grounding. These works emphasize the need for high-precision, context-aware annotation, but do not cover multilabel identity categories across many classes.

Gaps and Our Contributions. Current identity-focused corpora are typically single-label, English-centric, and restricted to broad political categories or explicitly stated group mentions. They rarely offer the fine-grained, multilabel annotations needed to capture the complexity of real-world political speech, and they do not cover non-Latin scripts or languages such as Hebrew. We introduce the first publicly available Hebrew dataset for social identity detection, comprising 5,536 sentences from politicians’ Facebook posts annotated with twelve distinct identities—each grounded in survey-measured salience and expert-defined categories. By providing multilabel annotations across a rich set of ideologically, religiously, and socio-economically relevant identities, our resource enables more nuanced analyses of how political actors deploy identity language than has previously been possible in any non-English setting.

3 Annotating Social Identities

3.1 Panel Survey

In order to choose social identities for annotation, we utilized a combination of experts and survey instruments. We utilized a representative 12-wave panel survey of the Jewish population in Israel (N = 1,769), conducted between January 2019 and April 2021 (ANONYMIZED REFERENCE). [2] This time period contains four Israeli elections held in quick

[2] Similar to other panel surveys in which local sample vendors were unable to include re-interview samples of Palestinian citizens in sufficient numbers (e.g., (Gidron et al., 2022), the panel survey did not include this population, reflecting a known deficiency in Israel’s survey sampling market.

succession, providing a unique opportunity to investigate political dynamics in a condensed time frame. In 12 survey waves, respondents were asked to select up to three identities they identify with, out of a list of 28. These 28 identities were chosen by a panel of experts in Israeli politics as reflecting a broad spectrum of prevalent social identities, spanning ideological, value-based, religious, national, socio-demographic, and economic dimensions (e.g., Conservative, Leftist, Nationalist, LGBTQ+. See Appendix A for a full list). For inclusion in the textual annotation, we selected the 12 identities that emerged as most salient in the panel survey, each consistently surpassing a 5% selection threshold in the first five waves. These identities are Rightist, Leftist, Conservative, Liberal, Socially-oriented, Capitalist, Zionist, Palestinian, Honest, Security-oriented, Ultra-orthodox, and Democrat.[3]

3.2 Annotation Scheme

Based on this selection, we developed an annotation scheme for the expression of social identities in text. Identities were annotated only when referenced in a positive manner; negative or oppositional mentions were excluded. Each sentence was annotated using a multilabel scheme across the twelve selected identities.

Below is a summary of the identity definitions used in the annotation scheme, accompanied by abridged examples translated from Hebrew (full definitions appear in appendix B).

Liberal: Advocacy for civil rights, pluralism, separation of religion and state, freedom of religion, protection of minority or LGBTQ+ rights, and support for the judicial system.

(1) *Protecting personal freedom and the right of every individual to be who they are and live as they choose.*

Conservative: Endorsement of opposition to change, support for the integration of religion and state, promotion of anti-liberal values, and advocating for the reduction of judicial authority.

(2) *The appointment of conservative and nationalist judges is the most significant factor in changing reality.*

Democrat: Emphasis on democratic values and procedures, such as fair elections, the rule of law, and institutional checks and balances. Includes references to the defense of democracy as a political principle.

(3) *The existence of a democratic, state-based regime in the country; the guarantee of the supremacy of the law.*

Leftist: Support for left-wing parties or policies, including dovish security positions, opposition to settlement construction, and criticism of right-wing actors or policies, when tied to a clear ideological stance. [4]

(4) *We believe that the evacuation of the territories occupied in the Six-Day War is a national necessity of the utmost urgency.*

Rightist: Support for right-wing parties or policies, including hawkish security positions, Greater Israel ideology, and criticism of left-wing actors or policies, when tied to a clear ideological stance.

(5) *We have an opportunity to form a fully right-wing government, so that we can pursue an unapologetic policy regarding Israeli settlements in Judea, Samaria, and the Jordan Valley.*

Capitalist: Support for free markets, deregulation, private enterprise, reduced government involvement in the economy, and growth-oriented economic policies that avoid redistributive or welfare-based framing.

(6) *The state should avoid regulatory intervention in the business sector as much as possible*

Socially-oriented: Support for social justice and welfare-oriented policies, including references to poverty, job security, government-funded education, healthcare, housing, and accessibility.

(7) *Masses of unemployed, sick, or needy individuals do not place their trust in the free market, but in the state - in its institutions, its elected officials, and its public servants.*

[3] The 5% criterion was jointly applied to identities that respondents identify with and identities they disapproved of in a separate survey item.

[4] Note that in Israel, both Leftist and Rightist identities primarily reflect individuals' positions on resolving the Israeli-Palestinian conflict, rather than economic issues.

Zionist: Affirmation of Zionist symbols and values, such as Jewish immigration to Israel, national pride, unity, and collective sacrifice (e.g., references to Memorial Day or the national anthem).[5]

(8) *It is a great source of pride to see our kindergarten children waving the flag with excitement and singing Independence Day songs.*

Security-oriented: Focus on national defense, military strength, borders, and threats to the internal or external security of the state.

(9) *Only targeting terrorist organizations will create deterrence, protect the security of Israeli citizens, and strengthen their resilience.*

Honest: References related to corruption and investigations involving public officials, honesty, and ethical conduct.

(10) *In two months, we will lead Israel off the path of corruption and onto a new path.*

Palestinians and Arab Citizens of Israel: Statements regarding policy issues, worldviews, and ideologies related to Palestinians and Arab-Israelis.

(11) *We will continue to fight for Arab local authorities and against budgetary discrimination of the Arab society.*

Ultra-orthodox: References to the Jewish ultra-Orthodox lifestyle, including the education system, gender segregation, and exemption from military service, as well as mentions of ultra-Orthodox political parties and leadership.

(12) *It is gratifying to see how the Shas movement succeeds in uniting communities through faith in God, the legacy of Rabbi Ovadia Yosef, and shared values.*

[5]This definition draws on the Declaration of Independence, which defines Israel as the homeland of all Jews while committing to full equality for all citizens and minority groups.

4 Dataset

To sample sentences for annotation, we compiled a corpus of Facebook posts by Israeli members of parliament, political parties, and viable party candidates, posted in the equivalent timeframe of the survey (Dec. 2018 to Apr. 2021). This corpus comprises 64,174 posts containing 375,718 sentences. Our annotated dataset contains 5,536 Hebrew-language sentences sampled from the Facebook posts corpus. Each sentence is annotated with 12 binary annotations according to whether it expresses a given identity. The dataset is split into a training set (70%), validation set (15%) and test set (15%).

4.1 Inter-coder Reliability

The sentences were annotated by two of the authors. We evaluate inter-coder reliability on a sample of 304 sentences. The results, given in Table 1, indicate a mean Cohen's Kappa of 0.77. Disagreements were resolved via consultation with all authors.

Category	Percent Agreement	Cohen's Kappa
Conservative	96.7%	0.705
Rightist	94.4%	0.788
Democrat	91.4%	0.703
Honest	96.7%	0.782
Capitalist	98.4%	0.792
Ultra-Orthodox	98.7%	0.708
Socially-oriented	96.1%	0.841
Liberal	96.1%	0.841
Leftist	94.1%	0.761
Security-oriented	96.4%	0.736
Palestinian	98.4%	0.792
Zionist	96.1%	0.778
Mean	96.1%	0.769

Table 1: Inter-Annotator Agreement (IAA) scores

4.2 Descriptive statistics

The number of positive labels per identity ranges between 129 (Ultra-Orthodox) and 703 (Rightist), with a mean of 413 (Table 2). 37.5% of sentences express no identity, 41.1% express one identity and 21.4% express two or more identities (Table 3).

5 Modeling and Experiments

We model the task of identifying social identities from text using three types of training setups:

Multilabel encoder-models are fine-tuned to jointly learn the 12 identity labels. **Single-label encoder models** are fine-tuned separately for each label. For this training, we utilize the best performing base model in the multilabel encoder setup. Finally, **Seq2seq LLMs** are fine-tuned to generate

Category	Count
Rightist	703
Liberal	629
Socially-oriented	567
Democrat	562
Leftist	546
Zionist	368
Honest	357
Security-oriented	346
Conservative	297
Capitalist	230
Palestinian	225
Ultra-Orthodox	129
Mean	413.25

Table 2: Number of positive instances

# Positives	Row Count	Percent
0	2,078	37.54%
1	2,275	41.09%
2	919	16.6%
3	218	3.94%
4	39	0.7%
5	6	0.11%
6	1	0.02%

Table 3: Number of Positive Labels per Sentence

a comma-separated list of all applicable Hebrew language labels for the given sentence.

We examine a set of encoder models and LLM decoders in the 2B-9B parameter ranges. For encoders, we use the multilingual mBERT (Devlin et al., 2019), and Hebrew-targeted encoders AlephBERT (Seker et al., 2022), HeRO (Hebron et al., 2023) and the base and large variants of DictaBERT (Shmidman et al., 2023). All encoder models are trained for up to 10 epochs with three learning rate (1e-5, 3e-5, 5e-5), and three type of loss (default, positive weight and focal loss), choosing the best performing checkpoint on the validation set.

For decoder LLMs, we use the multilingual Gemma 2 in the 2B and 9B variants (Team et al., 2024), the multilingual Qwen3-8B (Team, 2025), and the Hebrew-targeted DictaLM2.0 (Shmidman et al., 2024). All decoders are fine-tuned with QLORA (Dettmers et al., 2023) for up to 5 epochs with two learning rates (1e-5, 1e-4).

Table 4 reports test set results for trained models. We observe that the most performant are seq2seq models, achieving the highest per-label F1 scores on all but one label. DictaLM2.0, specifically, achieves the best macro P/R/F1 scores, and best per-label scores on 8/12 identities. The performance of this model, continuously pre-trained from Mistral 7B on 100B Hebrew tokens, underscores the dependence identity detection on language specific cultural and political context.

We also note that two of the multilabel encoders, $DictaBERT_{Large}$ and $DictaBERT_{Base}$, perform better on average than the single-label encoders finetuned for each identity separately using $DictaBERT_{Large}$ as a base model. $DictaBERT_{Large}$ achieved an F_1 score of 0.678 in a multi-label setting, compared to a mean F_1 of 0.659 when combining the single-label models. This indicates the benefit of joint learning and the interrelated nature of identities in this task.

5.1 Generalization to Parliamentary Speech

We examine whether the resulting classifier generalizes well to the parliamentary speech of the Israeli parliament (Knesset) floor by utilizing the IsraParlTweet dataset (Mor-Lan et al., 2024). We subset speeches from the equivalent time period of the Facebook data, classify all sentences with the best performing DictaLM2.0 model, and sample 500 of the sentences for human annotation. We perform a precision-oriented test by oversampling positive predictions for each identity. The results show a macro F_1 score of 0.72, on par with the Facebook test set, indicating the generalization ability of the model to Knesset data (for full results, see Table 5).

6 Results

We utilize the three identity data sources with overlapping time frames: our corpus of Israeli politician Facebook posts; the subset of parliamentary speeches from (Mor-Lan et al., 2024) in the same time period; and the panel survey capturing the public's identity choices. The first two sources are classified using the best performing DictaLM2.0 model. Utilizing these three data sources allows for a unique exploration of the dynamics of social identities on different platforms of elite discourse (Facebook and Knesset speeches) and the public (panel survey). We examine the following aspects: the popularity of identities; temporal trends; the correlation and bundling together of identities; and gender differences in identities.

Popularity. In Figure 1, we compare the normalized share of each identity. The identities *Socially-oriented*, *Rightist* and *Democrat* are generally pop-

	Macro-averaged metrics			Per-label F_1 scores											
Model	P	R	F_1	Security-oriented	Capitalist	Conserva-tive	Democrat	Ultra-Orthodox	Socially-oriented	Liberal	Palestinian	Leftist	Rightist	Honest	Zionist
Decoder-only															
DictaLM2.0	**0.740**	**0.751**	**0.743**	**0.705**	**0.805**	0.675	0.754	0.653	**0.723**	0.724	**0.852**	**0.750**	**0.765**	**0.816**	**0.700**
Gemma-9B	0.717	0.698	0.705	0.645	0.759	**0.684**	0.753	0.591	0.662	**0.749**	0.758	0.671	0.737	0.804	0.650
Gemma-2B	0.620	0.631	0.624	0.560	0.723	0.575	**0.757**	0.385	0.634	0.673	0.716	0.609	0.615	0.680	0.557
Qwen-8B	0.665	0.463	0.542	0.605	0.508	0.308	0.700	0.410	0.541	0.568	0.625	0.516	0.515	0.650	0.554
Multilabel encoders															
Dictabert-L	0.677	0.680	0.678	0.660	0.667	0.561	0.750	**0.667**	0.688	0.689	0.831	0.588	0.670	0.716	0.645
Dictabert-B	0.629	0.710	0.664	0.667	0.667	0.511	0.753	0.533	0.659	0.697	0.806	0.663	0.657	0.750	0.602
AlephBERT	0.628	0.692	0.657	0.673	0.675	0.526	0.738	0.604	0.627	0.664	0.783	0.620	0.641	0.718	0.618
HeRo	0.608	0.693	0.647	0.667	0.575	0.587	0.730	0.528	0.648	0.642	0.776	0.659	0.638	0.720	0.589
mBERT	0.573	0.553	0.552	0.483	0.467	0.527	0.708	0.356	0.568	0.635	0.632	0.531	0.502	0.653	0.566
Single-label encoders															
Dictabert-Large	0.643	0.689	0.659	0.672	0.659	0.529	0.731	0.508	0.708	0.694	0.783	0.615	0.636	0.721	0.652

Table 4: Macro-averaged precision (P), recall (R), F_1, and per-label F_1 for all models.

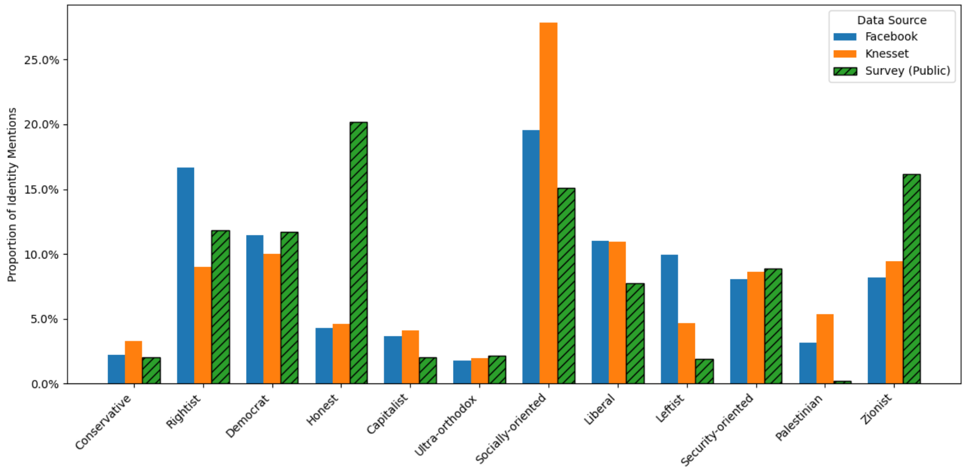

Figure 1: Normalized share of identities

ular across the data sources. Interestingly, the *Leftist* identity is relatively less popular. While many identities receive similar proportions among the three sources, several exceptions stand out. Identities *Honest* and *Zionist* are significantly more popular among the public, whereas *Socially-oriented* is significantly more popular in parliament.

Examining correlations between the ranks of identities in each data source, we see that while the popularity ordering of identities is strongly correlated between the Facebook and Knesset data (ρ=0.87), survey identity rankings are moderately correlated with Facebook (ρ=0.48) and Knesset (ρ=0.46) data. For all ranks, see Figure 10.

Temporal trends. We examine whether identity-related discourse tends to increase before elections. Facebook posts are the most suitable data source for this analysis, as the survey panel includes a limited number of waves and Knesset meetings are in recess near election periods. We plotted the average number of identity mentions per sentence per week in the Facebook data (see Figure 2). The results show considerable fluctuation, ranging from 0.38 to 0.87 identity mentions per sentence. Notably, identity-related discourse peaks around three of the four election dates, highlighting the salience of identities during electoral competition.

Figure 3 plots the share of six identities over time for both the Facebook and survey samples (see Figure 11 for all 12). We see that the identities that peak during or near election dates are *Rightist*, *Leftist* and *Democrat*. However, as previously mentioned, *Leftist* holds a very small share in both Facebook and the survey, and thus drives a smaller part of the overall temporal change. A notable gap between elites and the public emerges for three identities. While the public decreasingly identifies with the *Socially-oriented* identity, it becomes significantly more prominent among politicians after the 3rd elections held in March 2020. On the other hand, the identities *Honest* and *Democrat* become more popular among the public after the 3rd election, but less popular in elite discourse. These findings demonstrate the potential of HebID to provide

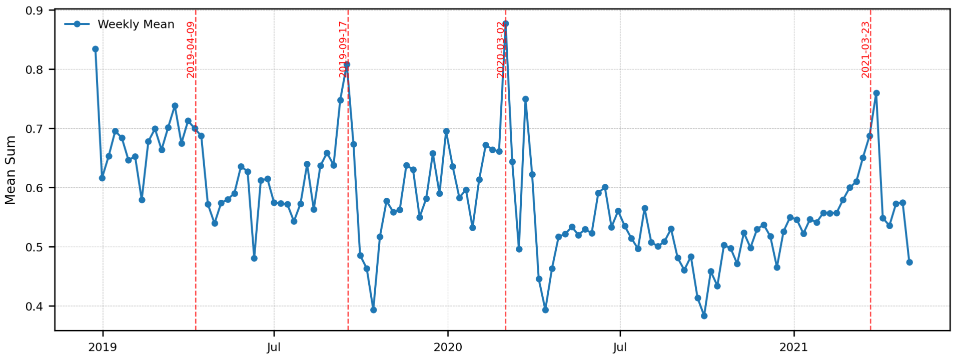

Figure 2: Temporal Trend

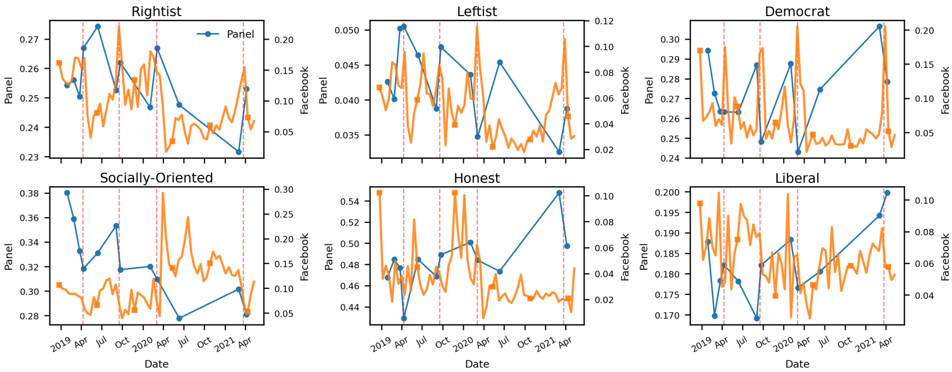

Figure 3: Per Identity Trends - Facebook and Survey

valuable insights into how identity-driven agendas shift in the lead-up to elections.

Identity bundles. To examine which identities tend to go together, we perform a factor analysis on the mean levels of each identity per speaker/respondent (Figure 4). In the three sources, we see a primary dimension dividing identities into two broad groups, a left-wing group (*Leftist*, *Democrat*, *Honest*, *Liberal*, *Palestinian*) and a right-wing group (*Rightist*, *Conservative*, *Zionist*, *Security-oriented*, *Capitalist*, *Ultra-orthodox*). Other factors in the Facebook and Knesset data appear to reflect political sub-dimensions of identity pairs that are more tightly coupled, such as *Conservative* and *Rightist*, *Zionist* and *Security-oriented*, *Liberal* and *Palestinian*, and *Honest* and *Democrat*. In the survey data, factors beyond the first dimension appear less structured compared to those that arise in the Facebook and Knesset datasets.

Accordingly, examination of the correlations between identities (after aggregating into speakers/respondents) shows that the survey exhibits a mean absolute correlation coefficient of 0.159, weaker than the Facebook sample (0.235) and Knesset data (0.215). For full correlation matrices, see the appendix.

Gender differences. Do men and women differ in terms of social identities? We first examine gender differences in identity discourse by calculating gender differences in the total number of identities expressed. For each speaker/respondent, we calculate the mean number of identities expressed in a sentence or survey response, and then aggregate by gender. We see that in all data sources, women express more identities than men. However, the gap is largest in the Knesset data (0.07) and is not statistically significant in the survey data.

We then examine the gender difference per identity by subtracting the share of each identity among women from that of men, in all data sources. We see that some identities, such as *Rightist*, *Security-oriented*, *Capitalist*, and *Ultra-orthodox* lean towards men, whereas *Socially-oriented* significantly lean towards women, on all platforms. While the gender gaps on Facebook and the Knesset are generally in the same direction, several of the survey gender gaps arise in different directions. Thus, *Honest* in the survey leans significantly towards women, whereas in the Facebook and Knesset data it leans slightly towards men.

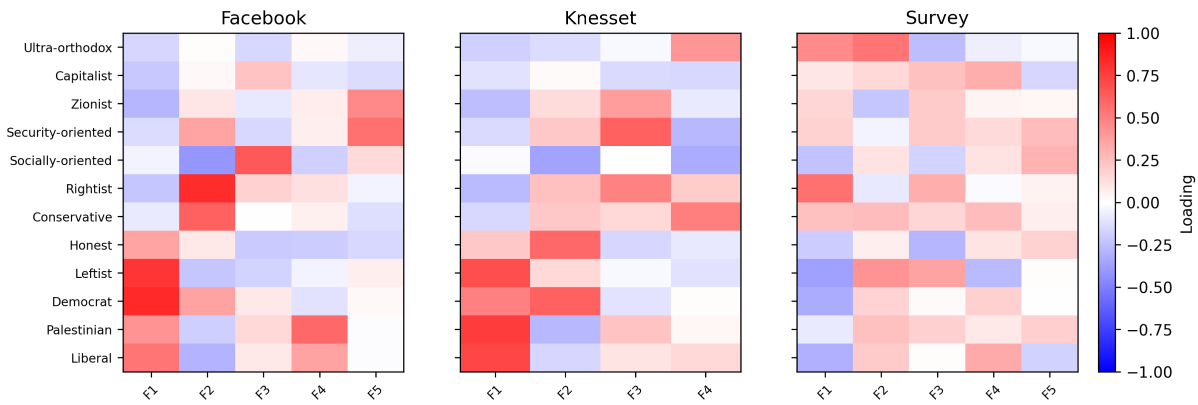

Figure 4: Factor analysis

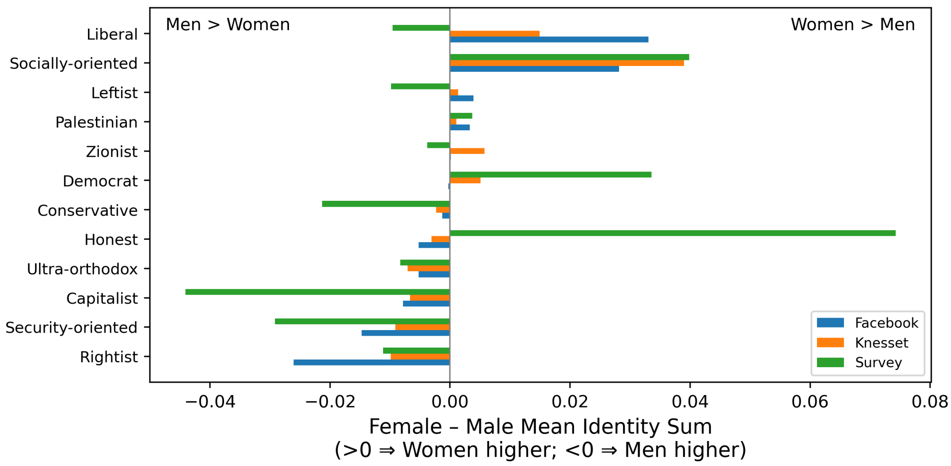

Figure 5: Gender Difference per Identity

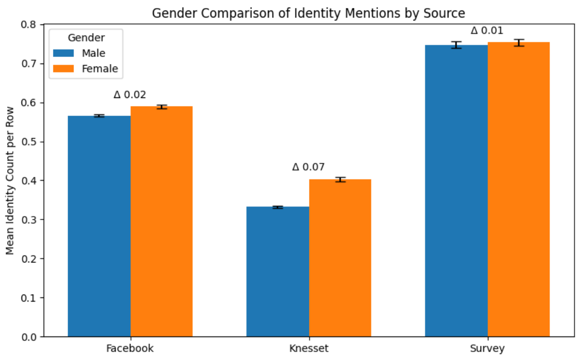

Figure 6: Gender Difference in Identity Mentions

7 Conclusion

In this work, we have introduced HebID, the first publicly available multilabel Hebrew corpus for fine-grained social identity detection in political text, grounded in large-scale survey evidence and expert consultation. Our dataset of 5,536 sentences from Israeli politicians' Facebook posts is annotated for twelve empirically salient identities, ensuring both linguistic and sociological validity. We benchmarked a range of models—multilabel and single-label encoders as well as 2B–9B-parameter seq2seq LLMs—and demonstrated that Hebrew-tuned seq2seq models (e.g., DictaLM2.0) achieve the highest macro-F1 (0.743), significantly outperforming encoder-only baselines. We further validated cross-genre robustness by applying our best model to 500 Knesset speech excerpts, matching social-media performance (macro-F1 = 0.72).

By linking three complementary sources — Facebook posts, parliamentary speeches, and a national survey, we revealed systematic differences in identity prevalence, temporal surges around elections, bundles of co-occurring identities, and gendered patterns of identity expression. The HebID annotation scheme and corpus, together with our code and models, provides a comprehensive framework for future research on identity discourse in non-English political contexts and paves the way for comparative studies across political systems.

Limitations

While HebID represents a significant step forward in Hebrew political text analysis, several limitations should be acknowledged:

- **Temporal and platform scope:** Identity discourse is dynamic and may have evolved beyond the period captured in our data. Additionally, other platforms—such as Twitter and news media—are not represented, leaving out potentially important dimensions of identity expression and change over time.
- **Survey population:** The panel survey sampled only Jewish citizens of Israel. Identities and salience patterns may differ among non-Jewish citizens, Palestinians, or immigrant communities.
- **Annotation granularity:** Although multilabel, our scheme relies on twelve categories selected via a 5% survey threshold. Less frequent but potentially important identities were excluded, and negative or critical references to an identity are not captured.
- **Model biases:** Our classifiers inherit biases present in both the training data and the pretrained language models (e.g., under- or over-representation of certain groups). Performance may degrade on dialectal text, informal registers, or in hostile political discourse.
- **Cross-genre validation:** The Knesset evaluation, while indicative of robustness, is based on a limited sample of 500 human-annotated sentences drawn from Knesset Plenum protocols. A broader evaluation across other legislative bodies or timeframes is needed to fully assess generalisation.
- **Methodological considerations:** Survey responses and text analysis offer complementary but distinct measurement approaches. Survey data provide insights into self-reported identity salience, while text analysis reveals patterns of expressed identity discourse. These methodological differences suggest caution when comparing panel survey results with computational text analysis findings.

Future work should extend HebID to additional platforms and populations, refine the annotation taxonomy to include emerging identities, and explore methods to mitigate model bias in identity detection tasks.

References

Dallas Card, Amber Boydstun, Justin Gross, Philip Resnik, and Noah Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of ACL-IJCNLP 2015*, pages 438–444.

Dallas Card, Kevin H. Chang, ..., and Philip Resnik. 2022. A century of rhetoric: Shifting sentiment on immigration in u.s. congressional speeches. *Proceedings of the National Academy of Sciences*, 119(14):e2117533119.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. Qlora: Efficient finetuning of quantized llms. *Preprint*, arXiv:2305.14314.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.

Noam Gidron, Lior Sheffer, and Guy Mor. 2022. The israel polarization panel dataset, 2019–2021. *Electoral Studies*, 80:102512.

David Hebron, Avi Shmidman, and Moshe Koppel. 2023. Hero: A hebrew roberta model for hebrew nlp tasks. *Preprint*, arXiv:2309.12345.

Pere-Lluís Huguet Cabot, David Abadi, Agneta Fischer, and Ekaterina Shutova. 2021. Us vs. them: A dataset of populist attitudes, news bias and emotions. In *Proceedings of the 16th Conference of the European Chapter of the ACL (EACL)*, pages 1921–1945.

Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *Proceedings of the 52nd Annual Meeting of the ACL*, pages 1113–1122.

Hauke Licht and Ronja Sczepanski. 2024. Who are they talking about? detecting mentions of social groups in political texts with supervised learning. Technical report, Working Paper.

Julia Mendelsohn, Ronan Le Bras, Yejin Choi, and Maarten Sap. 2023. From dogwhistles to bullhorns: Unveiling coded rhetoric with language models. In *Proceedings of ACL 2023*, pages 15162–15180.

Swaroop Mishra, Grzegorz Boratyn, and Martin Roemmele. 2023. Zero-shot ideological scoring of political text with large language models. *Transactions of the ACL*, 11:1234–1248.

Guy Mor-Lan, Effi Levi, Tamir Sheafer, and Shaul R. Shenhav. 2024. IsraParlTweet: The israeli parliamentary and Twitter resource. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 9372–9381, Torino, Italia. ELRA and ICCL.

Ines Rehbein and Josef Ruppenhofer. 2022. Who's in, who's out? predicting the inclusiveness or exclusiveness of personal pronouns in parliamentary debates. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5849–5858, Marseille, France. European Language Resources Association.

Amit Seker, Elron Bandel, Dan Bareket, Idan Brusilovsky, Refael Greenfeld, and Reut Tsarfaty. 2022. Alephbert: Language model pre-training and evaluation from sub-word to sentence level. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 46–56.

Avi Shmidman, Shaltiel Shmidman, and Moshe Koppel. 2024. Adapting llms to hebrew: Unveiling dictalm 2.0 with enhanced vocabulary and instruction capabilities. *Preprint*, arXiv:2407.07080.

Shaltiel Shmidman, Avi Shmidman, and Moshe Koppel. 2023. Dictabert: A state-of-the-art bert suite for modern hebrew. *Preprint*, arXiv:2308.16687.

Henri Tajfel and John C. Turner. 1979a. An integrative theory of intergroup conflict. In William G. Austin and Stephen Worchel, editors, *The Social Psychology of Intergroup Relations*, pages 33–47. Brooks/Cole, Monterey, CA.

Henri Tajfel and John C. Turner. 1979b. An integrative theory of intergroup conflict. In William G. Austin and Stephen Worchel, editors, *The Social Psychology of Intergroup Relations*, pages 33–47. Brooks/Cole, Monterey, CA.

Gemma Team, Morgane Riviere, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, Johan Ferret, Peter Liu, Pouya Tafti, Abe Friesen, Michelle Casbon, Sabela Ramos, Ravin Kumar, Charline Le Lan, Sammy Jerome, and 179 others. 2024. Gemma 2: Improving open language models at a practical size. *Preprint*, arXiv:2408.00118.

Qwen Team. 2025. Qwen3.

Michael Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of EMNLP 2006*, pages 327–335.

Sergio E. Zanotto, Qi Yu, Miriam Butt, and Diego Frassinelli. 2024. GRIT: A dataset of group reference recognition in italian. In *Proceedings of LREC-COLING 2024*, pages 7963–7970.

A Full list of social identities in Survey

1. Ashkenazi (Jewish of European descent)
2. Capitalist
3. Conservative
4. Democrat
5. A man of faith
6. Honest
7. Humanist
8. Israeli
9. Jewish
10. Leftist
11. LGBTQ+
12. Liberal
13. Lower class
14. Man
15. Middle class
16. Mizrahi (Jewish of Middle Eastern/North African descent)
17. Nationalist
18. Palestinian
19. Religious
20. Rightist
21. Secular
22. Security-oriented
23. Socialist
24. Socially-oriented
25. Ultra-Orthodox (Haredi)
26. Upper class
27. Woman
28. Zionist

B Identity Definitions: Codebook

Dear Coder,
Below are twelve identity definitions for classification. Please follow the following guidelines:

1. Classify Based on Definitions. Assign identity categories only to statements that align with the definitions provided.

2. Multiple Classifications. If a statement may fit more than one identity category. Please assign it to all relevant identity categories accordingly.

3. Positive Associations Only. Please note that the identity categories reflect a positive association. Do not classify statements that oppose an identity (e.g., criticism of capitalism, liberalism, Ultraorthodox) under that category.

4. The Speaker Identity Is Irrelevant. Classification is based on the content of the statement, not who says it (e.g., a non-Palestinian can make a statement classified under "Palestinians and Arab Citizens of Israel" if they present a Palestinian perspective).

5. Not all statements will be classified. Some statements may not fit any of the twelve identity categories mentioned in the codebook. In such cases, do not force a classification.

Identity definitions:

Capitalist. Support for private enterprises, free markets and economics (including the stock market), wealth accumulation, trade-related topics, deregulation policies, tax reductions, limiting government involvement in the economy or social services. General references to economic growth, market openness, or business encouragement—without mention of state intervention—will be labeled under this identity.
Socially-oriented (Hevrati). Support for social justice and welfare-oriented policies, including reference to poverty in Israel, job security, healthcare, government-funded education, social organizations, aid initiatives (e.g., food donations), social policy (e.g., housing assistance, support for disadvantaged populations), social security benefits, infrastructure investment, prioritizing local business over international ones, and environmental policy.
Conservative. Endorsement of opposition to change, support for the integration of religion and state, preservation of traditional values, promotion of anti-liberal values, and advocating for the reduction of judicial authority. Note: This identity includes reference that endorse Jewish traditions and opposition to anti-religious coercion.
Liberal. Advocacy for equality, human and civil rights, pluralism, separation of religion and state, freedom of religion, the promotion of universal values, protection of minority or LGBTQ+ rights, and support for the judicial system.
Note that:

1. References relating to inequality in the burden of civic duties will be labeled under this identity.

2. This identity does not include references to the rule of law.

Democrat. Emphasis on democratic values and procedures, such as fair elections, the rule of law, institutional checks and balances ("rules of the game"), and the defense of democracy as a political principle. Includes explicit references to democracy or portraying Israel as a democratic state.
Honest (Yeshar Derech). References related to corruption and investigations involving public officials, honesty, integrity, and ethical conduct in public service. Includes mentions of improper immunity, criticism of self-serving behavior by public officials, and praise for individuals acting in public interest. Note that neutral factual statements (e.g., "Netanyahu's trial begins tomorrow in Jerusalem") and descriptive mentions of corruption without a clear perspective will not be classified under this identity.
Leftist. Support for left-wing parties or policies, including dovish security policies, willingness to make territorial compromise, opposition to settlement construction, criticism of right-wing actors or policies, when tied to a clear ideological stance.
Note that:

1. Identification with left-wing parties or groups is labeled as Leftist.

2. This identity focuses on conflict-related positions, not economic or social issues (which are coded separately, e.g., Liberal or Socially-oriented).

3. References to "peace" alone do not constitute a leftist identity without additional ideological context.

4. Centrist parties opposing a side are labeled according to the target of opposition (e.g., “Blue and White opposes Likud” = Leftist; “Blue and White opposes Labor” = Rightist).

5. Positions regarding state-religion relations are not part of ideological identities (e.g., referring to religious coercion would be considered liberal, not Leftist). Moreover, attacking ultra-Orthodox parties does not constitute a Leftist identity.

Rightist. Support for right-wing parties or policies, including hawkish security positions, Greater Israel ideology, criticism of left-wing actors or policies, when tied to a clear ideological stance.
Note that:

1. Identification with right-wing parties or groups is labeled as Rightist.

2. This identity focuses on conflict-related positions, not economic or social issues (which are coded separately, e.g., Conservative or Capitalist).

3. References to Israel as a “Jewish and democratic state” do not alone indicate Rightist identity.

4. Centrist parties opposing a side are labeled according to the target of opposition (e.g., “Blue and White opposes Labor” = Rightist; “Blue and White opposes Likud” = Leftist;).

5. Positions regarding state-religion relations are not part of ideological identities (e.g., referring to greater congruence between state and religious laws would be considered Conservative not Rightist).

Palestinians and Arab Citizens of Israel. Statements regarding policy issues, worldviews, and ideologies related to Palestinians and Arab-Israelis, including references to Palestinian and Arab-Israeli culture, statements and interviews by Palestinian and Arab-Israeli public leaders related to Palestinians and Arab-Israelis, as well as policy matters concerning these groups (not just security issues).
Note that:

1. References that do not reflect a Palestinian perspective (e.g., discussions on settlements or military actions against Palestinian organizations), will not be labeled under this identity.

2. The speaker does not have to be Palestinian but must present events or impacts from a Palestinian perspective.

3. Positive references for public figures representing these communities will be included under this identity.

4. This identity is distinct from the Leftist identity. References can be to both or either, depending on context.

Security-Oriented (Bitchonist). References to national security issues, military strength, security capabilities, threats to internal or external safety, defense agencies (IDF, Shin Bet, Mossad, etc.), borders protection, the state’s ability to protect its citizens and security challenges.
Note that:

1. This identity includes references that relate to the notion of protecting all citizens (e.g., Jewish, Arab-Israeli) from violence or terror.

2. This identity excludes general or symbolic mentions of soldiers unrelated to defense (e.g., prayers, greetings, daily life).

3. This identity includes critiques of external actors (e.g., referring to the Palestinian Authority as a terrorists funding organization) when framed in terms of national security.

4. “No peace without security” statements fall under this identity.

Ultraorthodox (Haredi). References to Jewish ultraorthodox (Haredi) lifestyle, including the Haredian education system, gender segregation in the public sphere, exemption from military service for yeshiva students, charitable activities within the Haredi community, the Haredi religious (Torah) world and tradition preservation, Haredian parties (Shas, United Torah Judaism, Agudat Yisrael), and Haredi rabbinic leadership.
Note that:

1. The mere mentioning of Rabies is not enough. Statements must carry a positive tone.

2. General religious expressions (e.g., quoting a verse, mentioning Jewish holidays) will not be labeled under this identity unless clearly framed within Haredian context or authority.

3. References to the integration of Haredim into broader Israeli society, or criticism of the Haredi community, will not be labeled under this identity.

Zionist. Affirmation of Zionist symbols and values, such as Jewish immigration to Israel (Aliyah), connections between Israel and the Jewish diaspora, national pride, IDF enlistment, national unity, symbolic expressions such as the positive references to the Israeli flag, national anthem and collective sacrifice (e.g., references to Memorial Day, families of fallen soldiers). Also includes references relating to Israel's struggle against antisemitism and BDS. Note that:

1. This definition draws on the Declaration of Independence, which defines Israel as the homeland of all Jews while committing to full equality for all its citizens, including minority groups.

2. This identity includes expressions of dedication to Israel's well-being or future ("to serve the country," "for the future of the country").

3. This identity does not include general mentions of antisemitism without connection to Israel, references to soldiers unrelated to enlistment or national service.

C Fine-tuning setup

All model fine-tuning is performed on an 80GB A100 Nvidia GPU, using huggingface transformers.

For decoder fine-tuning, the separator "### Answer:" is used to separate the input sentences from the output. The labels of the input and separator are loss-masked.

All experiments utilize AdamW optimizer with a linear scheduler. Default values of hyperparamters are used everywhere except for learning rate (for encoders and for decoder LLMs) and loss type (for encoder models).

Decoder fine-tuning uses QLORA with 4bit quanitzation. LORA settings are rank of 256, and alpha value of 512. LORA layers are attached to all linear levels in the decoder models.

Each hyper-parameter configuration was trained once.

D Generalization to Parliamentary Speech

Table 5 shows the full results of the 500 items sampled from parliamentary Knesset speeches for cross-genre generalization test. The predictions are produced by the best performing training checkpoint (DictaLM2.0 fine-tune).

Class	Precision	Recall	F1-score	Support
Conservative	0.33	0.82	0.47	17
Rightist	0.66	0.74	0.70	47
Democrat	0.76	0.69	0.72	74
Honest	0.72	0.81	0.76	47
Capitalist	0.68	0.87	0.76	31
Ultra-orthodox	0.66	0.84	0.74	32
Socially-oriented	0.73	0.72	0.73	72
Liberal	0.60	0.60	0.60	63
Leftist	0.66	0.77	0.71	53
Security-oriented	0.78	0.75	0.77	48
Palestinian	0.91	0.86	0.88	69
Zionist	0.76	0.81	0.78	42
Micro avg	0.70	0.76	0.73	595
Macro avg	0.69	0.77	0.72	595
Weighted avg	0.72	0.76	0.73	595
Samples avg	0.62	0.64	0.62	595

Table 5: Knesset sample results – DictaLM2.0 fine-tuned checkpoint

E Data Release

The annotated data is released under cc-by-4.0 license. The data is publicly available on github at https://github.com/guymorlan/pdd/.

F Annotation

All data has been annotated by two authors of the paper. The authors have not received direct compensation. The two annotators are Hebrew speaking women living in Israel.

G Model Sizes

Model	Parameters
mBERT (bert-base-multilingual-cased)	110 M
AlephBERT	110 M
HeRo	125 M
DictaBERT–base	184 M
DictaBERT–large	340 M
Gemma–2B	2 B
Gemma–9B	9 B
Qwen3–8B	8.2 B
DictaLM 2.0	7 B

Table 6: Model sizes (number of parameters) for all models used in this paper.

H Preprocessing

Sentence segmentation was performed using the Stanza package.

I Additional Results

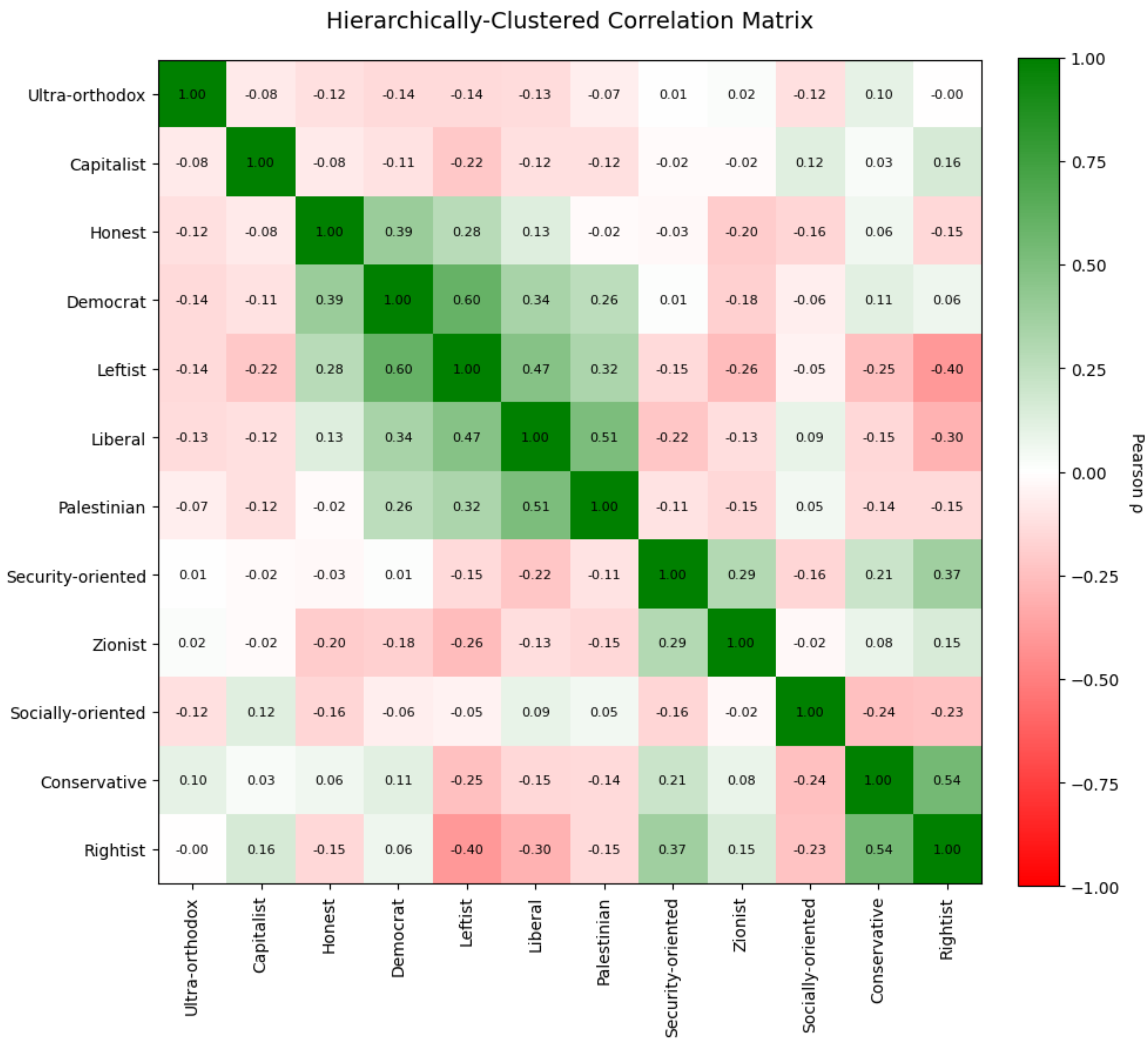

Figure 7: Identity Correlation Matrix - Facebook

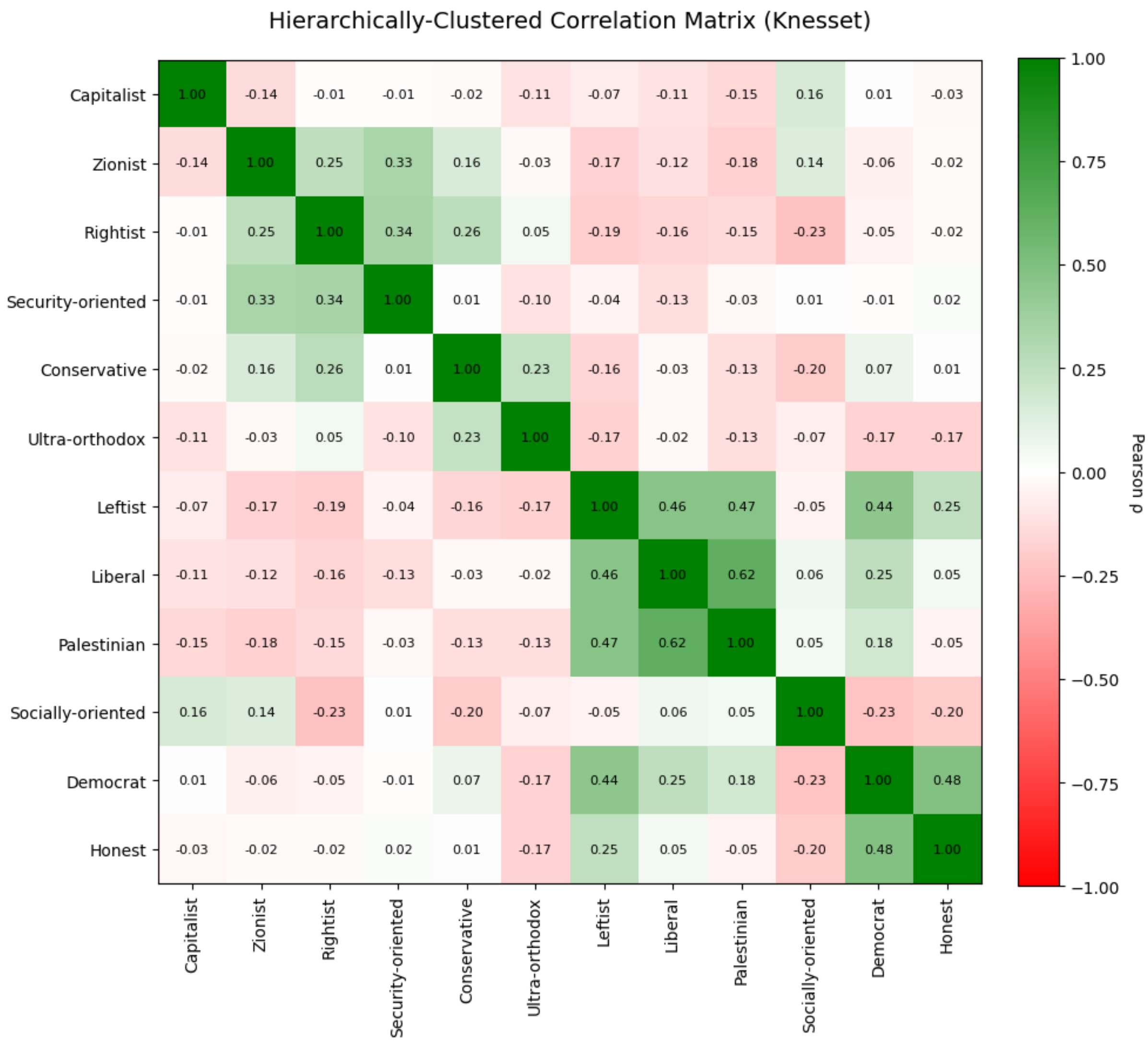

Figure 8: Identity Correlation Matrix - Knesset

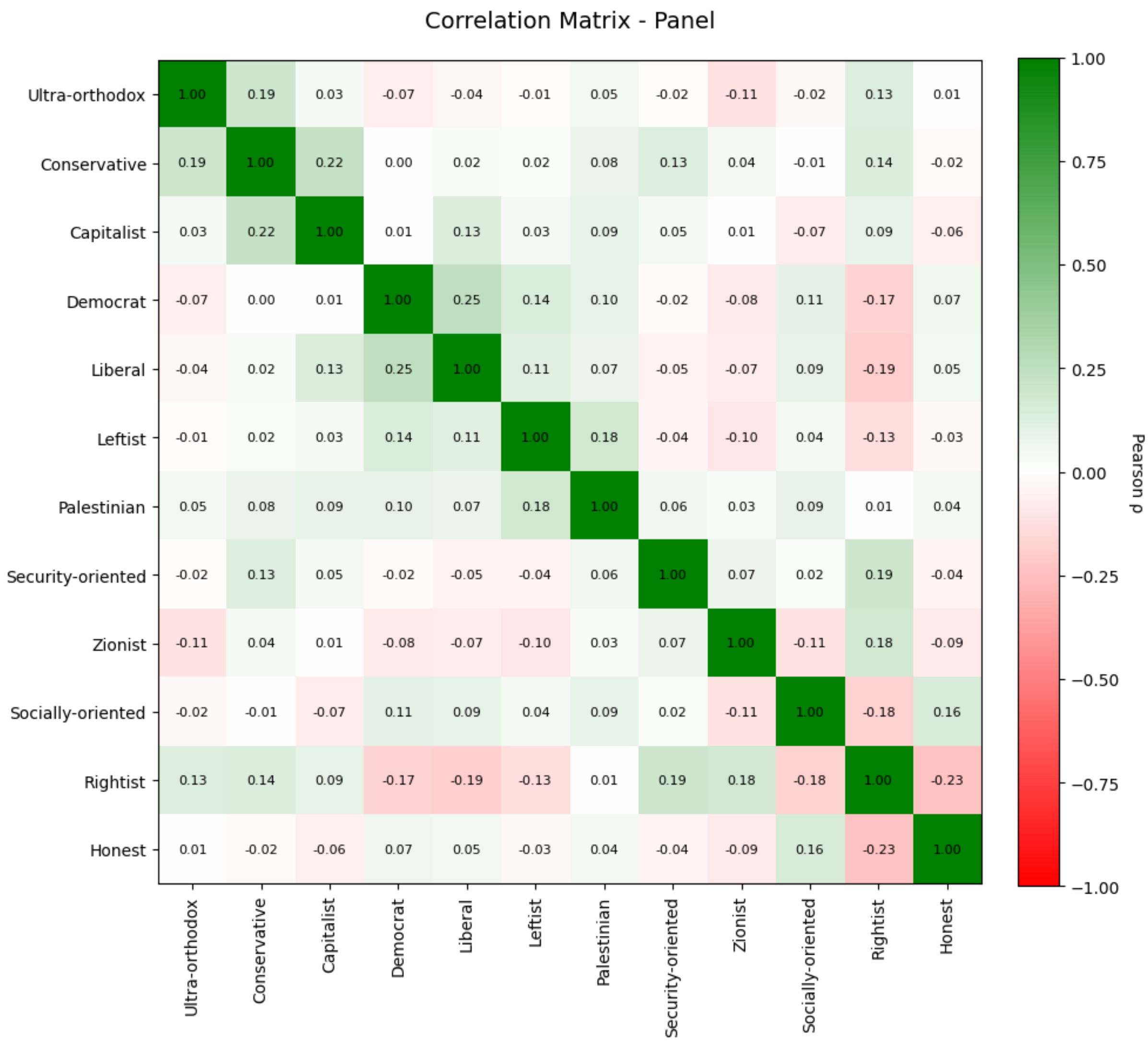

Figure 9: Identity Correlation Matrix - Survey

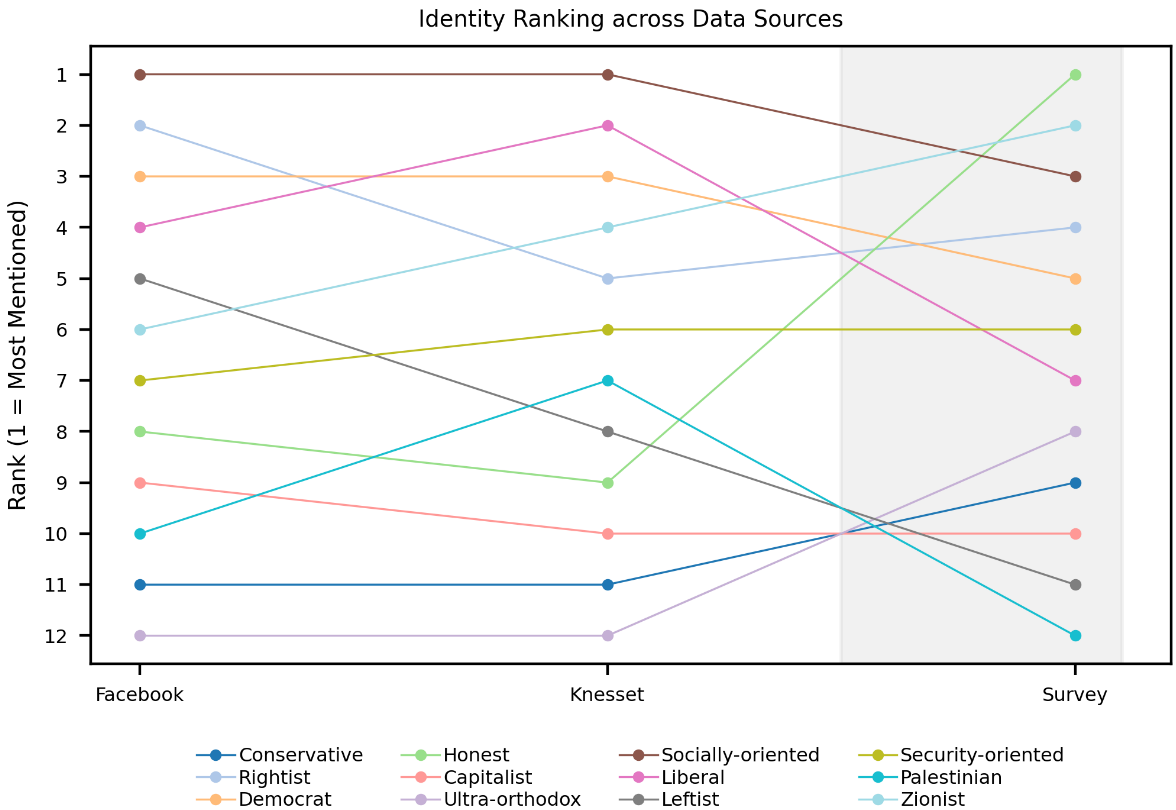

Figure 10: Identity Ranking

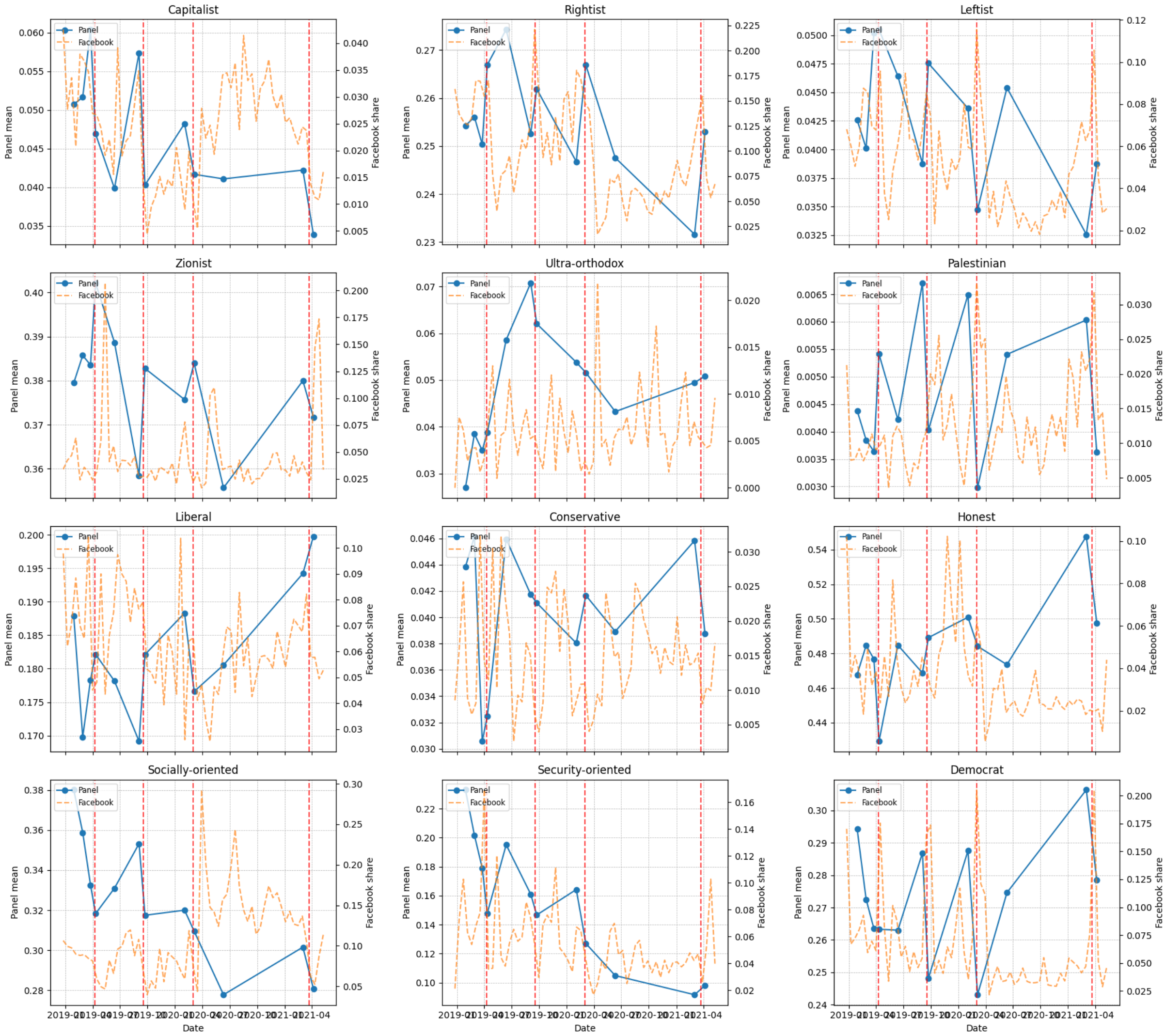

Figure 11: Per-identity time-trends on Facebook (biweekly mean) and survey