Political Analysis (2025), **00**, 1–8
doi:10.1017/pan.2025.7

PA

LETTER

Mapping (A)Ideology: A Taxonomy of European Parties Using Generative LLMs as Zero-Shot Learners

Riccardo Di Leo[1], Chen Zeng[2], Elias Dinas[3] and Reda Tamtam[4]

[1]Research Fellow, Department of Political and Social Sciences, European University Institute, Fiesole, Italy; [2]Research Associate, Department of Political and Social Sciences, European University Institute, Fiesole, Italy; PhD Student, Department of European and International Studies, King's College London, London, UK; [3]Swiss Chair in Federalism, Democracy and International Governance, Department of Political and Social Sciences, European University Institute, Fiesole, Italy; [4]Research Associate, Department of Political and Social Sciences, European University Institute, Fiesole, Italy; Graduate Student, Department of Politics, Princeton University, Princeton, NJ, USA

Corresponding author: Riccardo Di Leo; Email: riccardo.dileo@eui.eu

(Received 12 July 2024; revised 24 November 2024; accepted 21 January 2025)

Abstract

We perform the first mapping of the ideological positions of European parties using generative Artificial Intelligence (AI) as a "zero-shot" learner. We ask OpenAI's Generative Pre-trained Transformer 3.5 (GPT-3.5) to identify the more "right-wing" option across all possible duplets of European parties at a given point in time, solely based on their names and country of origin, and combine this information via a Bradley–Terry model to create an ideological ranking. A cross-validation employing widely-used expert-, manifesto- and poll-based estimates reveals that the ideological scores produced by Large Language Models (LLMs) closely map those obtained through the expert-based evaluation, *i.e.*, CHES. Given the high cost of scaling parties via trained coders, and the scarcity of expert data before the 1990s, our finding that generative AI produces estimates of comparable quality to CHES supports its usage in political science on the grounds of replicability, agility, and affordability.

Keywords: ideology scores; computational methods; expert opinion; matched pairs; text-as-data

Edited by: Daniel J. Hopkins and Brandon M. Stewart

The study of party ideology is arguably one of the most compelling examples of the importance of measurement in social sciences. Testing and refining theories of party competition and representation would have been impossible without viable comparative measures of parties' policy positions. The numerous technical solutions proposed in the literature tend to fall within two camps: manifesto- and expert-based measures. Whilst party manifestos allow researchers to study a broad range of years and countries, ideological measures derived from this corpus do not always perform highly in terms of face (Laver, Benoit, and Garry 2003) and convergent (Dinas and Gemenis 2010) validity. Expert surveys fare better in both dimensions (Hooghe *et al.* 2010), but are costly to gather, *de facto* unavailable before the 1990s, and—by definition—involve a high degree of human arbitrariness, threatening the replicability of their measures.[1]

In this paper, we ask whether recent advancements in Large Language Models (LLMs) may allow social scientists to bridge the two historical approaches to classification. Can we reliably recur to machine learning to obtain real-time, low-cost measures of parties' ideological positions that match the validity of experts, while retaining the geographical and chronological breadth of manifestos?

[1]Yet, see Mikhaylov, Laver, and Benoit (2012) for evidence of low cross-coder reliability in manifestos.

LLMs such as the proprietary GPT language model, the open-weight Llama, and the open-source Qwen represent a relatively novel alternative to supervised learning models in text analysis.[2] Self-supervised tasks continuously train an artificial neural network over a virtually infinite corpus of digital and physical media, and "the target prediction is provided within the data itself, rather than hand-labeled by a researcher" (Ornstein, Blasingame, and Truscott 2025, 5). This arguably represents a major advantage for social scientists, as gathering trained experts or untrained crowds and asking them to annotate a large amount of data is a costly and time-consuming endeavor (Ornstein *et al.* 2025), prone to conformability issues (Mikhaylov *et al.* 2012).

Artificial Intelligence (AI)-generated data presents several advantages in terms of both *reproducibility* and *agility* of the data generation process, as defined by Benoit *et al.* (2016). On the one hand, the use of a unique "seed" allows replicators to reconstruct the AI-generated data, with relatively little time and effort: a task hardly attainable in coder and expert surveys, due to the difficulty of recontacting samples, and to the volatility and context dependence of human responses (Sanders, Ulinich, and Schneier 2023). On the other hand, researchers using LLMs can easily adjust several parameters, including the degree of "randomness" allowed for the algorithm to complete the task. Furthermore, performing large-scale tasks via LLMs is a relatively affordable exercise.[3] Finding that generative AI produces estimates of comparable quality to those obtained via the commonly used alternatives—experts, manifestos, and/or opinion polls—would support its usage in social sciences on the grounds of replicability, agility, and affordability. Our aim is to verify whether this is the case.

In this paper, we perform, to the best of our knowledge, the first structural assessment of whether LLMs can correctly evaluate the left-right position of European parties as "zero-shot" learners, *that is*, without being provided any context regarding the task. We favor this methodology over a "few-shot" approach as the two have been shown to perform equally well in classification tasks (Le Mens and Gallego 2025; Ziems *et al.* 2024). In the latter, though, the researcher arbitrarily chooses what information is fed (or not) to the algorithm, and how so, *e.g.*, providing examples of "correct" classifications (Section G of the Supplementary Material), or explicit ideological stances (Le Mens and Gallego 2025). Arguably, the zero-shot approach removes this final layer of human intervention in the creation of the training set, increasing the reproducibility of the process.

We verify the convergent validity of the ideological positions estimated by OpenAI's Generative Pre-trained Transformer 3.5 (GPT-3.5)[4] by looking at whether AI-generated estimates fall within the range of those produced by expert surveys, party manifestos, and opinion polls. We ascertain that estimates produced by GPT-3.5 are remarkably close to those coming from the experts, slightly less proximate to the ideological scores assigned by voters, and even less so to those obtained from party manifestos. We then briefly illustrate a battery of tests, presented in greater detail in the Supplementary Material, aimed at verifying the robustness of our methodology and illustrating its limitations.

AI applications in political science have been numerous (for a review, see Ornstein *et al.* 2025): from the annotation and interpretation of political texts in "few-shot" learning contexts (*e.g.*, Gilardi, Alizadeh, and Kubli 2023), to the creation of credible "synthetic" survey samples (*e.g.*, Argyle *et al.* 2023). Still, only a few papers in political science have so far exploited GPT-3.5 as a "zero-shot" learner. Wu *et al.* (2023) showed ChatGPT can effectively rank US congressmen's ideologies through pairwise comparisons, while Bol and Bono (2024) found that GPT-4 performs well in the former dimension even when asked to position French political parties on the left-right axis, solely based on their names. We expand the latter study in both the geographical and time scope, employing the methodology developed in the former. More generally, our work contributes to a growing strand of research studying

[2] Although released as "open-source", Llama-3 fails to meet several of the Open Source Initiative (OSI) criteria, *e.g.*, model weights are available, but not its training data/code. For a discussion, see: spectrum.ieee.org. Qwen is released under the Apache License 2.0, approved by the Free Software Foundation and the OSI. All online resources were last accessed on November 24, 2024.

[3] A detailed breakdown of the costs is presented in Section L.2 of the Supplementary Material. Using GitHub Actions, we timestamped each API call, uploading online the timing of each iteration, its duration, and results.

[4] Ornstein *et al.* (2025) find that GPT-3.5 outperforms GPT-4 in "few-shot" prompting.

whether LLMs can "transform computational social science" by deepening our understanding of social phenomena and constructs, including political ideology (Ziems *et al.* 2024, 238).

1. Data and Methods

The pool of European party names presented, in their original language, to GPT-3.5, and the scores used to benchmark its estimates of parties' ideology, are gathered from three sources: (1) the *Manifesto* Project (CMP, Lehmann *et al.* 2024); (2) the Chapel Hill *Expert* Survey, combined with the Ray–Marks–Steenbergen dataset (CHES, Jolly *et al.* 2022; Ray 1999); (3) the harmonized True European *Voter* survey (TEV, Schmitt 2021).

We generate nine cross-country sub-samples of parties to capture the European arena at the time of each European Parliament (EP) election between 1979 and 2019. To increase comparability across sources, we retain in each sub-sample only parties having run—and having been evaluated—in an election taking place at most 4 years *prior* to the EP contest under scrutiny (for more information, see Section A of the Supplementary Material).

To estimate the LLM-based left-right position of a party in a given reference year, we implement a three-step procedure: (1) we produce all the possible pair-wise party duplets in the reference year, irrespective of the countries of origin; (2) we ask GPT-3.5 to identify the "more right-wing" option between the two parties in a duplet; (3) we apply a Bradley–Terry model to these duplets, generating an ideological ranking based on GPT-3.5's evaluations (hereafter, *GPT-BTm*) in which cross-party distances are interpretable (Loewen, Rubenson, and Spirling 2012). Pairwise comparisons have been widely employed as a classification method (*e.g.*, Hopkins and Noel 2022), and are preferable to rankings when utilizing GPT-3.5 (Wu *et al.* 2023).

We perform each classification exercise seven times and retain the modal answer for our analysis. Showing that responses to pairwise comparisons are highly consistent in repeated interactions—congruence is as high as 93.23%—provides further validation for our method (for a comparable exercise, see Wu *et al.* 2023).[5]

2. Results

In Figure 1, we explore the correlation between our LLM-generated ideological scaling of parties on the left-right scale (GPT-BTm) and the ones obtained via CMP *manifestos* (first row), CHES *experts* (second row), and TEV *voters* (third row). In each panel, we report the Pearson coefficient between GPT-BTm and the validation dataset under scrutiny, in a given reference year. GPT-BTm scores are highly correlated with experts' (ρ = 0.78 in 2009) and, to a lesser extent, voters'. They exhibit the lowest congruence with manifesto-based evaluations. The Pearson correlation coefficients do not vary significantly over time, especially when CHES is the benchmark. A more detailed inspection of the geographical heterogeneities sketched in Figure 1—alongside linguistic ones—reveals that GPT-BTm scores are stable when evaluated against CHES or TEV, but less so when it comes to CMP: the correlation between the latter and GPT-3.5 is lower in Southern European and Slavic-speaking countries (Sections I.2 and I.3 of the Supplementary Material).

In the Supplementary Material, we perform a battery of tests to verify the robustness of our findings and identify the limitations of our approach. We explore uncertainty in GPT-3.5's classifications using log-probabilities and Bayesian inference in Section B of the Supplementary Material, finding that the LLM reports high levels of confidence.

We show that GPT-3.5 performs equally well when pooling all reference years together (Section C of the Supplementary Material), and when asked to estimate parties' positions after its training period ends (Section D of the Supplementary Material): GPT-3.5's ability to perform "out-of-sample" predictions

[5]The GPT-3.5 setup and the computation of GPT-BTm scores are outlined in Section L of the Supplementary Material, the full inference procedure in Section M of the Supplementary Material.

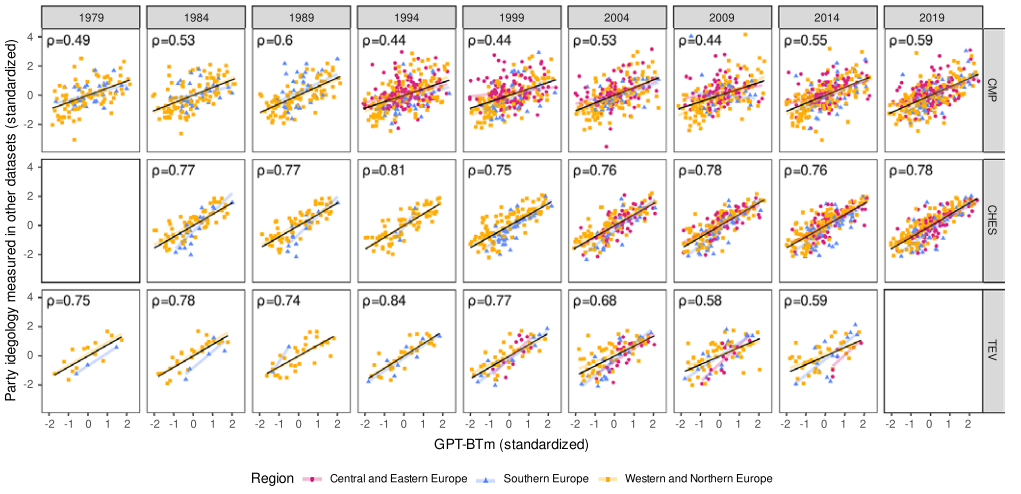

Figure 1. Benchmarking European parties' left-right positions according to GPT-3.5 against experts, manifestos, and opinion polls. *Note:* In each facet, we plot the left-right ideological positions of European parties obtained applying a Bradley–Terry model to the pairwise comparisons performed by GPT-3.5 (GPT-BTm), in a given reference year (indicated on top of the sub-figure). In each row we employ different validation datasets, namely: Chapel Hill Expert Survey (CHES), Comparative Manifesto Project (CMP), True European Voter (TEV). The varying number of observations across each panel and, in turn, the different number of GPT-BTm comparisons performed, reflect the different sampling criteria adopted by each data source, as outlined in Section A of the Supplementary Material.

would distinguish it from CMP/CHES as the sole "prospective" classifier, although a more thorough investigation appears necessary to back such conclusion. A related question is whether GPT-BTm may solve the issue of existing ideological scores registering low levels of convergent validity in longitudinal settings (Adams *et al.* 2019), despite being widely employed to proxy policy *shifts* across elections (*e.g.*, Adams and Somer-Topcu 2009). In Section E of the Supplementary Material, we show that GPT-3.5 does not improve on the other benchmarks in this respect.

The regression analysis in Section F of the Supplementary Material provides more direct evidence that, in a horse race against CMP, CHES is a more powerful and important predictor of GPT-BTm scores. Also, our zero-shot approach outperforms a four-shot one (Section G of the Supplementary Material).

In Section H of the Supplementary Material, we verify that the scores are insensitive to the model setup, namely, to the framing of the task (Section H.1 of the Supplementary Material), the ordering of the party duplets (Section H.2 of the Supplementary Material), and choice of the "temperature" parameter (Section H.3 of the Supplementary Material), determining the LLM's "leeway" in the output-generating process (Sanders *et al.* 2023). In Section I of the Supplementary Material, we address potential compositional effects: GPT-3.5's responses are insensitive to randomly cutting the sample of evaluated duplets, and to outliers (Section I.1 of the Supplementary Material).

We adopt alternative benchmarks for the LLM's ideological classification in Section J of the Supplementary Material. On the one hand, we find comparable results as for CHES and TEV when employing alternative experts' and voters' surveys (Section J.1 of the Supplementary Material) coming from the Comparative Study of Electoral System (CSES, 2024). On the other, as CMP estimates are sensitive to the methodology employed to process them (Dinas and Gemenis 2010), we test several alternatives to *rile*, namely, the "vanilla" method (Gabel and Huber 2000), log odds-ratios (Lowe *et al.* 2011), and the valence-based "regression" (Franzmann and Kaiser 2006). GPT-BTm correlates almost as highly with the last index as with experts' scores (Section J.2 of the Supplementary Material).

In Section K of the Supplementary Material, we replicate our exercise via different LLMs. We compare different versions of the LLMs developed by OpenAI, finding that GPT-4o outperforms GPT-3.5 and GPT-4o mini, yet at a higher cost per iteration (Section K.1 of the Supplementary Material). We then shift our attention to the open-weight LLM Llama-3.1 (Section K.2 of the Supplementary Material),

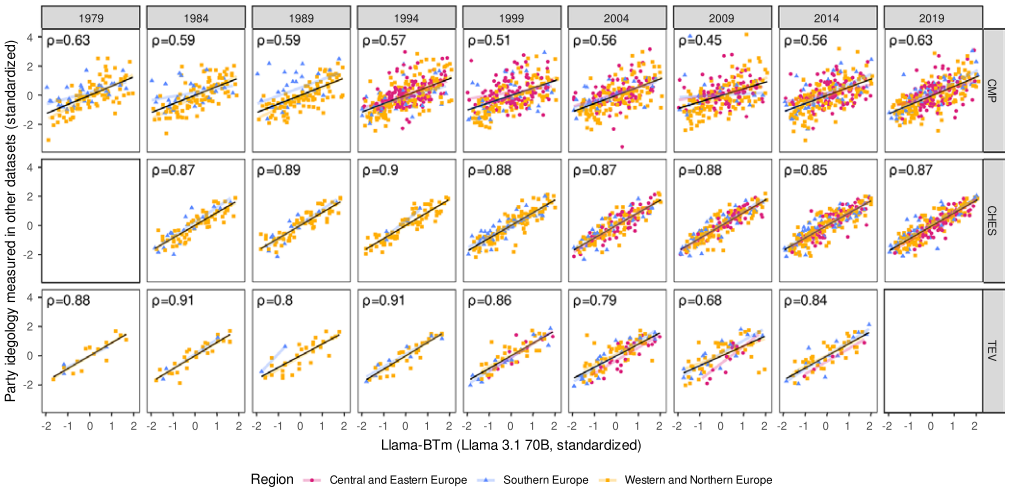

Figure 2. Benchmarking European parties' left-right positions according to Llama-3.1 70B against experts, manifestos and opinion polls.

Note: In each facet, we plot the left-right ideological positions of European parties obtained applying a Bradley-Terry model to the pairwise comparisons performed by Llama-3.1 70B (Llama-BTm), in a given reference year (indicated on top of the sub-figure). In each row we employ different validation datasets, namely: Chapel Hill Expert Survey (CHES), Comparative Manifesto Project (CMP), True European Voter (TEV). The varying number of observations across each panel and, in turn, the different number of Llama-BTm comparisons performed, reflect the different sampling criteria adopted by each data source, as outlined in Section A of the Supplementary Material.

showing that its 70B model outperforms GPT-3.5 in the classification task (Figure 2). We compare two versions of Llama trained on the same corpus, but released with different parameter sizes, to show that the results are stable across model updates, keeping the information set constant (Section K.2 of the Supplementary Material). Altogether, these tests reassure us about the robustness of our results to changes in the architecture underlying the LLMs, although different results may emerge from future releases.

In Section K.3 of the Supplementary Material, we show that the performance of the open-source LLM Qwen-2.5 32B rivals that of proprietary and open-weight LLMs. Other open-source models, instead, seem ill-suited for zero-shot ideological classification. Finally, in Section K.4 of the Supplementary Material, we systematically examine ten years of Common Crawl releases, a widely-employed training resource. The inspection suggests that the raw data underlying our benchmarks is either irretrievable or in a format that LLMs cannot parse, providing reassurance against the risk of contamination, *i.e.*, of LLM-produced scores "parroting" those assigned by experts/coders/voters.

3. Discussion

LLMs have proven to be able to convincingly navigate the US political arena (Argyle *et al.* 2023). Yet, arguably, multi-polar European party systems present a whole different set of challenges.

In our paper, we provided the first cross-country classification of European parties' ideological positions using LLMs. Our analysis improves our understanding of where to locate GPT-3.5 as an ideological coder within the range of measures available to social scientists. It reveals that the AI-based assessment of European parties' ideological positions is remarkably closer to those obtained from experts and crowds, than to those coming from analyzing the textual content of party manifestos.

Leaving aside the scholarly debate on the pros and cons of expert- and manifesto-based approaches to scaling ideology, finding, as we do, that estimates obtained via Generative AI are close to those provided by experts, and that this congruence holds over time, implies that AI can become a valid substitute of expert-based measures, with much wider spatial and temporal coverage. More generally, our findings

confirm GPT-3.5's ability as a "Zero-Shot" political learner (Wu *et al.* 2023), and support the growing consensus on LLMs' potential in augmenting the computational social sciences pipeline (Ornstein *et al.* 2025).

Our results are necessarily exploratory in nature, and some caveats are in order. LLMs are not *ex-ante* trained to perform "zero-shot" learning. They remain a conceptual "black box", where tracking the sources the deep neural network employs to form a decision is *de facto* impossible (Törnberg 2024). Necessarily, the quality of GPT-3.5's assessment depends upon the availability of a large quantity of training data (Bender *et al.* 2021): scholars wishing to apply this methodology to contexts where information is scarce—or questions of difficult interpretation—should consider LLMs' tendency to "hallucinate", *i.e.*, to report implausible answers with high levels of confidence, in these settings (Bang *et al.* 2023; Törnberg 2024). More generally, researchers should be wary of the limitations of LLMs as ideological classifiers. Two main questions emerging from our analysis deserve, in our view, a more thorough inspection. Why does GPT-3.5 fail to accurately capture ideology in a dynamic setting, not dissimilarly from existing measures? And why do some open-source LLMs perform poorly in a zero-shot setting?

On a related note, the increased use of LLMs by social scientists may dampen, in itself, the reliability of AI, raising the risk of self-contamination—*i.e.*, of extant evaluation exercises becoming themselves a "source" for subsequent ones (Aiyappa *et al.* 2023). Finally, even if, in a "zero-shot" context, the researcher does not exert a *direct* influence on the algorithm, the scaling produced by LLMs will nevertheless incorporate any *indirect* bias coming from the vast corpus of human-produced text on which the model is pre-trained (Caliskan, Bryson, and Narayanan 2017), posing non-trivial ethical concerns. As a result, the "ideology" measured by GPT-3.5 aligns more closely with the reputational, rather than policy-based, aspect of party positioning, which also resonates with how voters and experts may perceive ideology. The high correlation between the scores from GPT-3.5 and CHES suggests that, if anything, their assessments may be informed by similar biases.

In light of this discussion and of the several methodological alternatives available to researchers, two crucial questions remain: why use LLMs to scale parties' political positions? And how to interpret the ideological classifications LLMs produce? We second Ornstein *et al.* (2025, 19): finding that "despite its limitations, the GPT-3 approach yields estimates that correlate strongly with human-coded measures at a small fraction of the cost [. . .] has enormous practical implications". We argue that scholars can employ LLMs to accurately capture parties' political leaning, overcoming the methodological concerns raised by manifesto-based assessments, the low supply of experts in the field of inquiry, and the contextual biases of national opinion polls. The use of LLMs as ideological classifiers presents two important advantages: on the one hand, they can deliver rankings in real-time, rather than as a (lagged) by-product of an election. On the other, LLMs are affordable to all researchers, regardless of their affiliation or seniority. Both motivations, together with the high correlation with the classifications coming from experts, suggest that LLMs could provide an ideal tool for pilot studies, and/or in circumstances where up-to-date ideological scores are unavailable.

Acknowledgments. We are grateful to the anonymous reviewers and the editor for their valuable comments and input. We thank Cathrine Kjaer for the helpful feedback on a previous version of the article. This work has benefited from computation time provided by Princeton Research Computing.

Author Contributions. R.D.L. and C.Z. have contributed equally to this work.

Competing Interest. The authors declare none.

Funding Statement. This work was funded by the European Union (ERC, POSTNORM, 101088868). Views and opinions expressed are however those of the authors and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Data Availability Statement. Replication code for this article has been published in Code Ocean, a computational reproducibility platform that enables users to run the code, and can be viewed interactively at https://doi.org/10.24433/CO.5139195.v1. A preservation copy of the same code and data can also be accessed via Dataverse at https://doi.org/10.7910/DVN/SECNCZ

(Di Leo *et al.* 2025). Results obtained from OpenAI's GPT models are available on GitHub Actions at https://github.com/realczeng/postnorm_mapping_main. We timestamped each API call being performed, uploading online the timing of each iteration, its duration, and results.

Supplementary Material. For supplementary material accompanying this paper, please visit https://doi.org/10.1017/pan.2025.7.

References

Adams, J., L. Bernardi, L. Ezrow, O. B. Gordon, T.-P. Liu, and M. C. Phillips. 2019. "A Problem with Empirical Studies of Party Policy Shifts: Alternative Measures of Party Shifts Are Uncorrelated." *European Journal of Political Research* 58 (4): 1234–1244.

Adams, J., and Z. Somer-Topcu. 2009. "Policy Adjustment by Parties in Response to Rival Parties' Policy Shifts: Spatial Theory and the Dynamics of Party Competition in Twenty-Five Post-War Democracies." *British Journal of Political Science* 39 (4): 825–846.

Aiyappa, R., J. An, H. Kwak, and Y.-Y. Ahn. 2023. "Can We Trust the Evaluation on Chat GPT?". doi: https://doi.org/10.48550/arXiv.2303.12767. Pre-published.

Argyle, L. P., E. C. Busby, N. Fulda, J. R. Gubler, C. Rytting, and D. Wingate. 2023. "Out of One, Many: Using Language Models to Simulate Human Samples." *Political Analysis* 31 (3): 337–351.

Bang, Y., et al. 2023. "A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity." https://doi.org/10.48550/ARXIV.2302.04023. Pre-published.

Bender, E. M., T. Gebru, A. McMillan-Major, and S. 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*. Association for Computing Machinery, New York, NY, USA, 610–623. doi: https://doi.org/10.1145/3442188.3445922.

Benoit, K., D. Conway, B. E. Lauderdale, M. Laver, and S. Mikhaylov. 2016. "Crowd-Sourced Text Analysis: Reproducible and Agile Production of Political Data." *American Political Science Review* 110 (2): 278–295.

Bol, D., and P.-H. Bono. 2024. "Est-ce que ChatGPT est une Bonne Politologue?" Note de recherche, Élections européennes 2024, vague 3 de l'enquête électorale, note 5. Accessed February 25, 2025: https://www.sciencespo.fr/cevipof/sites/sciencespo.fr.cevipof/files/Noteelectionseuropeennes_DB%26PHB_chatGPT_mai2024_VF.pdf. Pre-published research note.

Caliskan, A., J. J. Bryson, and A. Narayanan. 2017. "Semantics Derived Automatically from Language Corpora Contain Human-like Biases." *Science* 356 (6334): 183–186.

The Comparative Study of Electoral Systems (CSES). 2024. "CSES Integrated Module Dataset (IMD) Version 4.0.0". GESIS Data Archive, Cologne. doi:10.4232/cses.imd.2024-02-27.

Di Leo, R., C. Zeng, E. Dinas, and R. Tamtam. 2025. "Replication Data for: Mapping (A)Ideology: A Taxonomy of European Parties Using Generative LLMs as Zero-Shot Learners." Harvard Dataverse, Version V1. https://doi.org/10.7910/DVN/SECNCZ.

Dinas, E., and K. Gemenis. 2010. "Measuring Parties' Ideological Positions With Manifesto Data: A Critical Evaluation of the Competing Methods." *Party Politics* 16 (4): 427–450.

Franzmann, S., and A. Kaiser. 2006. "Locating Political Parties in Policy Space: A Reanalysis of Party Manifesto Data." *Party Politics* 12 (2): 163–188.

Gabel, M. J., and J. D. Huber. 2000. "Putting Parties in Their Place: Inferring Party Left-Right Ideological Positions from Party Manifestos Data." *American Journal of Political Science* 44 (1): 94.

Gilardi, F., M. Alizadeh, and M. Kubli. 2023. "ChatGPT Outperforms Crowd Workers for Text-Annotation Tasks." *Proceedings of the National Academy of Sciences* 120 (30): e2305016120.

Hooghe, L., et al. 2010. "Reliability and Validity of the 2002 and 2006 Chapel Hill Expert Surveys on Party Positioning." *European Journal of Political Research* 49 (5): 687–703.

Hopkins, D. J., and H. Noel. 2022. "Trump and the Shifting Meaning of "Conservative": Using Activists' Pairwise Comparisons to Measure Politicians' Perceived Ideologies." *American Political Science Review* 116 (3): 1133–1140.

Jolly, S., et al. 2022. "Chapel Hill Expert Survey Trend File, 1999–2019." *Electoral Studies* 75: 102420.

Laver, M., K. Benoit, and J. Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97 (2): 311–331.

Le Mens, G., and A. Gallego. 2025. "Positioning Political Texts with Large Language Models by Asking and Averaging." *Political Analysis*. Published online: 1–9. doi:10.1017/pan.2024.29.

Lehmann, P., et al. 2024. "The Manifesto Project (MRG/CMP/MARPOR) Version V2024a." Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB), Göttingen: Institut für Demokratieforschung (IfDem). doi: 10.25522/manifesto.mpds.2024a.

Loewen, P. J., D. Rubenson, and A. Spirling. 2012. "Testing the Power of Arguments in Referendums: A Bradley–Terry Approach." *Electoral Studies* 31 (1): 212–221.

Lowe, W., K. Benoit, S. Mikhaylov, and M. Laver. 2011. "Scaling Policy Preferences from Coded Political Texts." *Legislative Studies Quarterly* 36 (1): 123–155.

Mikhaylov, S., M. Laver, and K. R. Benoit. 2012. "Coder Reliability and Misclassification in the Human Coding of Party Manifestos." *Political Analysis* 20 (1): 78–91.

Ornstein, J. T., E. N. Blasingame, and J. S. Truscott. 2025. "How to Train Your Stochastic Parrot: Large Language Models for Political Texts." *Political Science Research and Methods* 13 (2): 264–281.

Ray, L. 1999. "Measuring Party Orientations towards European Integration: Results from an Expert Survey: Research Note." *European Journal of Political Research* 36 (2): 283–306.

Sanders, N. E., A. Ulinich, and B. Schneier. 2023. "Demonstrations of the Potential of AI-based Political Issue Polling." *Harvard Data Science Review* 5 (4): 1–37.

Schmitt, H. 2021. "The True European Voter Version 1.0.0." GESIS Data Archive, Cologne. doi: 10.4232/1.13601.

Törnberg, P. 2024. "Large Language Models Outperform Expert Coders and Supervised Classifiers at Annotating Political Social Media Messages." *Social Science Computer Review*. Published online: 1–15. doi: 10.1177/08944393241286471.

Wu, P. Y., J. Nagler, J. A. Tucker, and S. Messing. 2023. "Large Language Models Can Be Used to Estimate the Latent Positions of Politicians." doi: 10.48550/arXiv.2303.12057. Pre-published.

Ziems, C., W. Held, O. Shaikh, J. Chen, Z. Zhang, and D. Yang. 2024. "Can Large Language Models Transform Computational Social Science?" *Computational Linguistics* 50 (1): 237–291.

Cite this article: Di Leo, R., Zeng, C., Dinas, E. and Tamtam, R., (2025). Mapping (A)Ideology: A Taxonomy of European Parties Using Generative LLMs as Zero-Shot Learners. *Political Analysis*, 1–8. https://doi.org/10.1017/pan.2025.7