Measuring the Quality of Answers in Political Q&As with Large Language Models

Online Supplementary Material

R. Michael Alvarez and Jacob Morrier

Division of the Humanities and Social Sciences

California Institute of Technology

November 2024

Contents

List of Figures

List of Tables

A Additional Figures and Tables

Table A1: Training Hyperparameters

Model	`multi-qa-mpnet-base-cos-v1`
Loss Function	Multiple Negatives Ranking Loss
Epochs	10
Batch Size	8
Optimizer	AdamW†
Learning Rate	2×10^{-5}†
Learning Rate Scheduler	Warm-up Linear†
Warm-up Steps	10,000†
Weight Decay	0.01†
Maximum Gradient Norm	1†

†Default Value

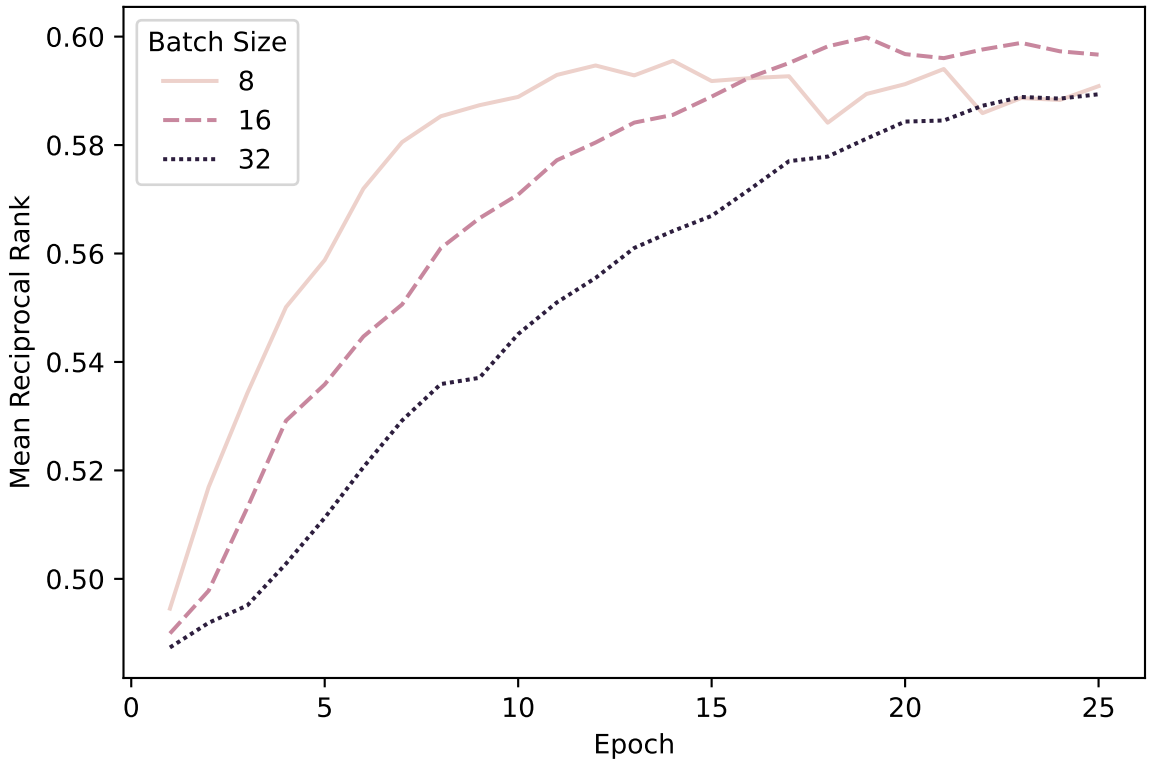


Figure A1: Mean Reciprocal Rank of the Validation Set by Batch Size and Epoch

Table A2: Descriptive Statistics of the Distribution of the Cosine Similarity Between Questions and Answers

Count	54,914
Mean	0.5387
Standard Deviation	0.1865
Minimum	−0.1625
First Quartile	0.4178
Median	0.5608
Third Quartile	0.6806
Maximum	0.9542

Table A3: Mean Reciprocal Rank of the Inference Set

Model Variant	Pre-Trained	Fine-Tuned	Fine-Tuned ("Reverse" Objective)
Question	0.1157	0.1260	**0.1276**
Answer	0.1040	**0.1292**	0.1281

Note: The first row shows how well model variants rank the possible questions related to each answer, while the second row shows how well model variants rank the possible answers for each question.

Table A4: Performance of the Fine-Tuned Model on the Inference Set

	@ 10	@ 25	@ 100
Precision	0.0256	0.0144	0.0054
Recall	0.2561	0.3588	0.5497
F-1 Score	0.0232	0.0138	0.0054

Note: Each column presents the performance metrics for some threshold, with all answers ranked above that threshold being predicted as correct. For example, the first column shows our model's precision, recall, and F-1 score when, for each question, the top 10 answers (ranked by their cosine similarity to the question's embedding) are predicted as correct. In this scenario, 2.56% of predictions are accurate, and 25.61% of correct answers are captured by these predictions.

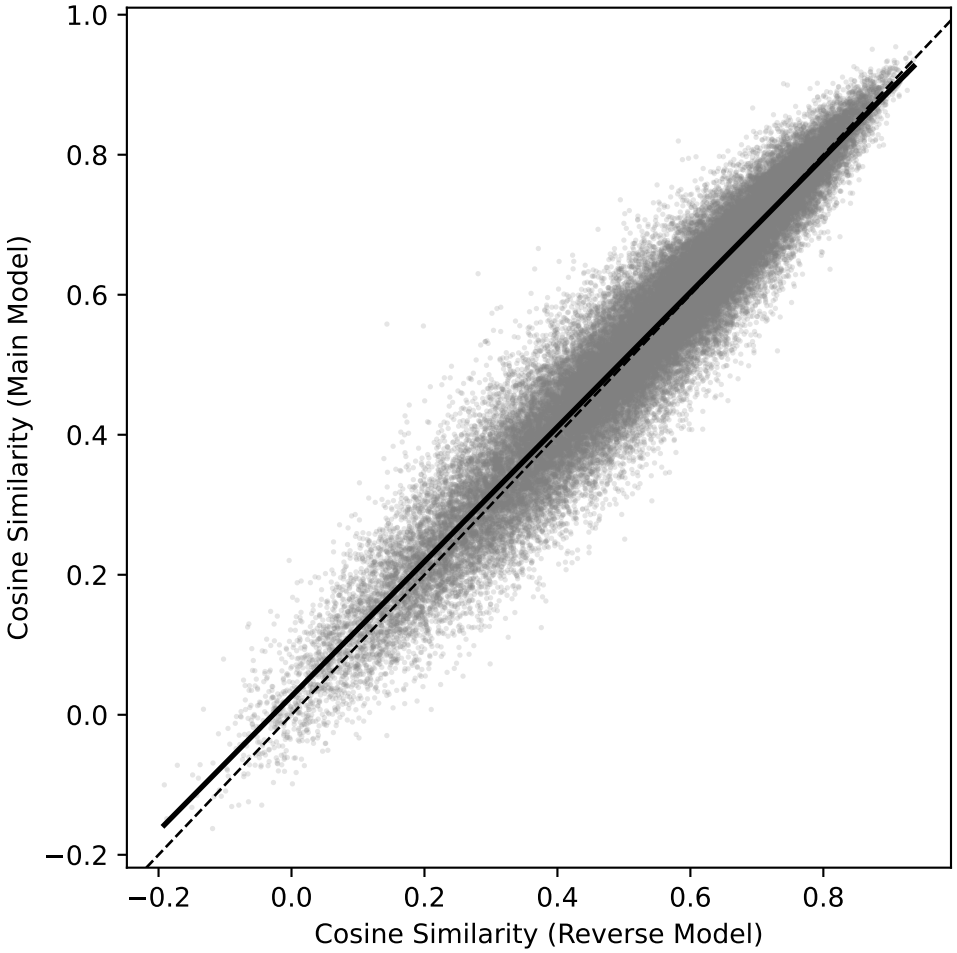


Figure A2: Comparison of Cosine Similarity Estimates Between Fine-Tuned Models with Main and Reverse Objectives

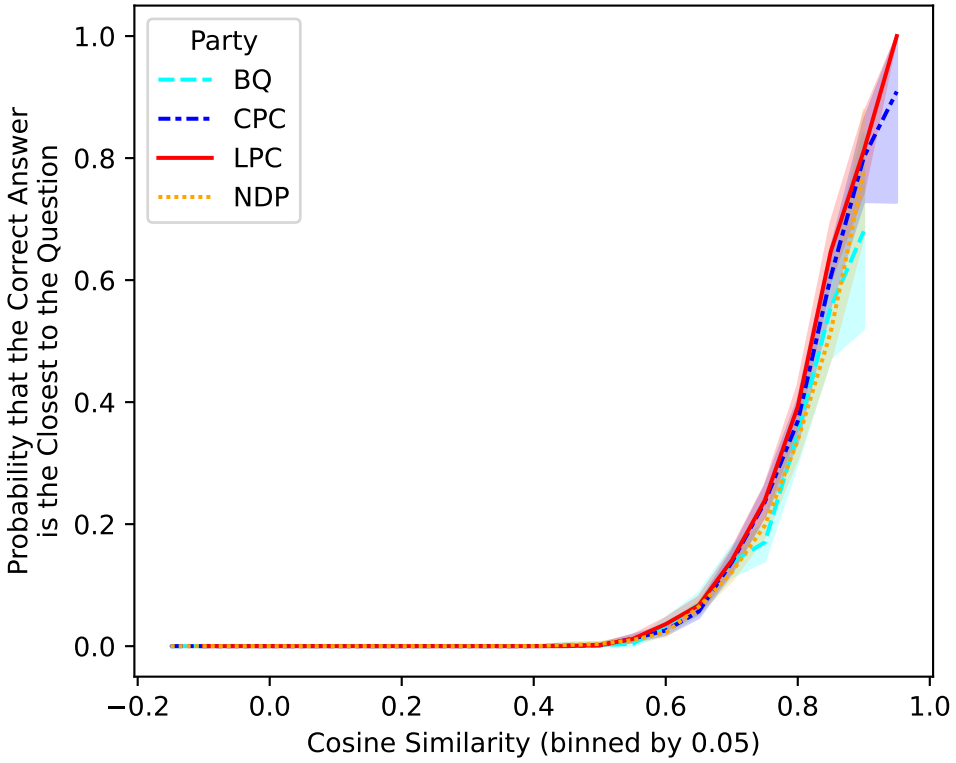


Figure A3: Probability that the Correct Answer is the Closest to the Question by Cosine Similarity Between Questions and Answers and Party

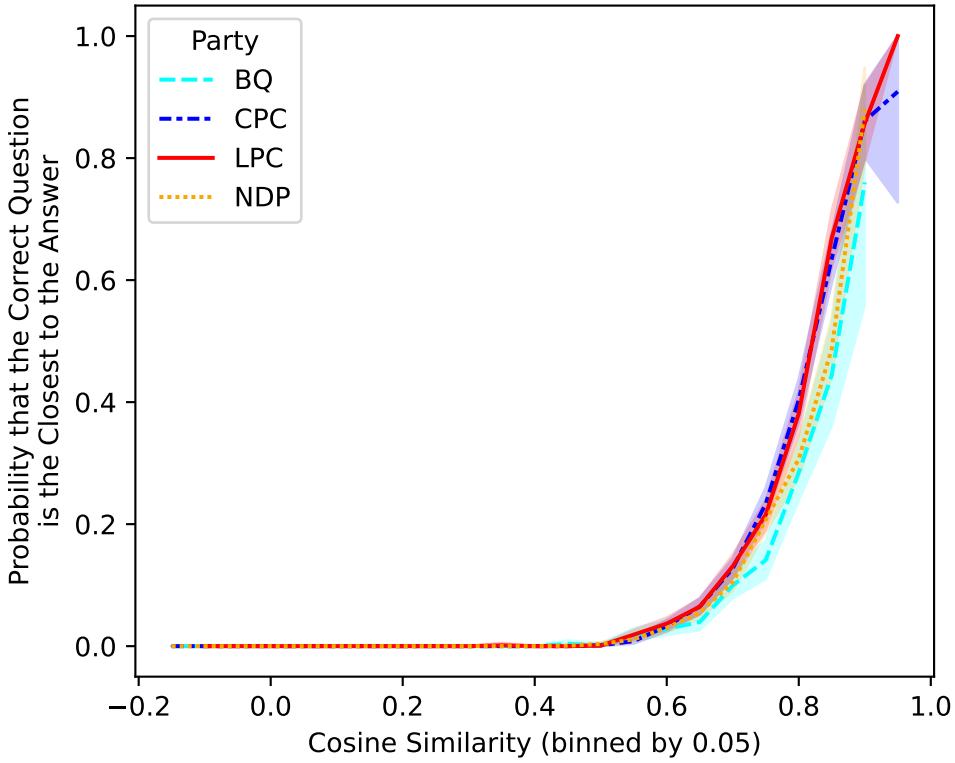


Figure A4: Probability that the Correct Question is the Closest to the Answer by Cosine Similarity Between Questions and Answers and Party

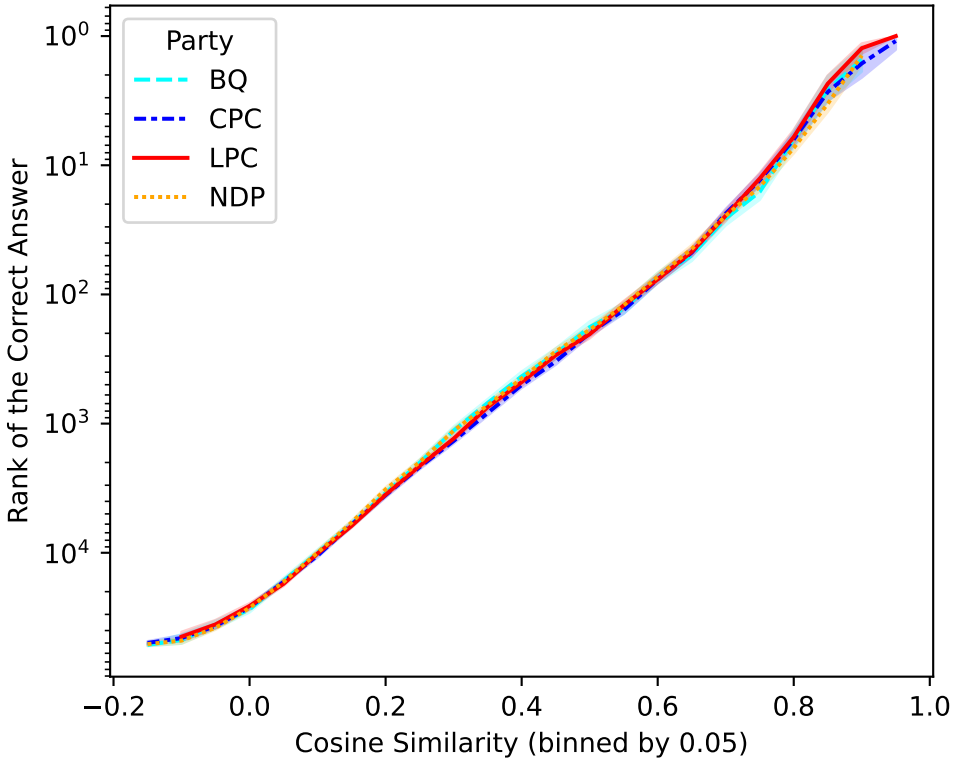


Figure A5: Rank of the Correct Answer by Cosine Similarity Between Questions and Answers and Party

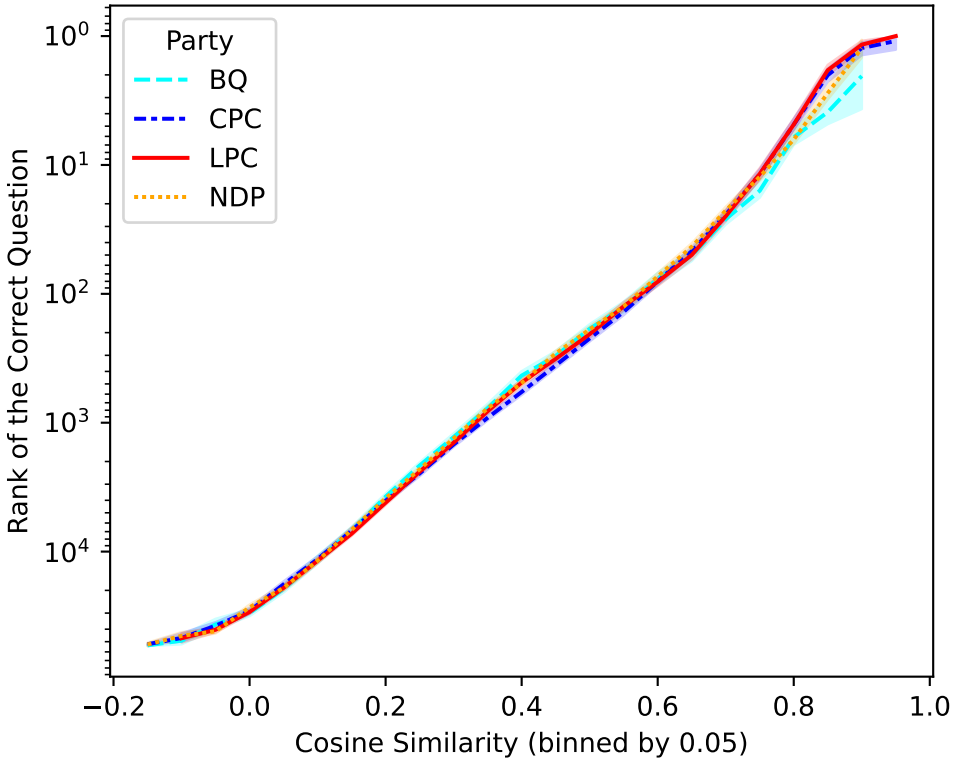


Figure A6: Rank of the Correct Question by Cosine Similarity Between Questions and Answers and Party

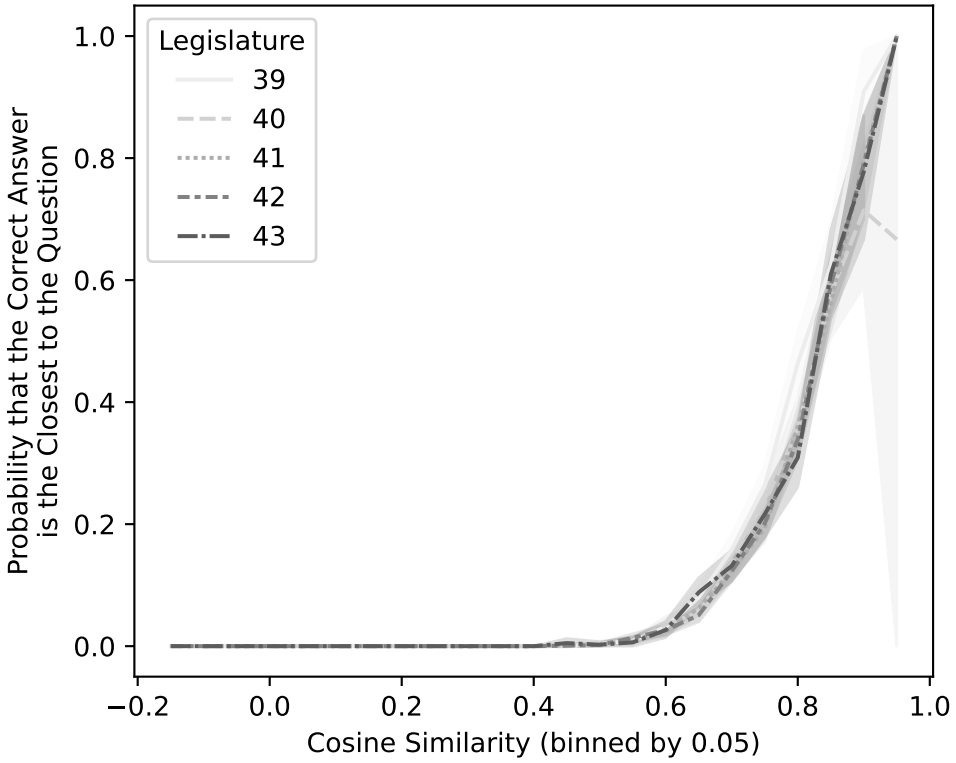


Figure A7: Probability that the Correct Answer is the Closest to the Question by Cosine Similarity Between Questions and Answers and Legislature

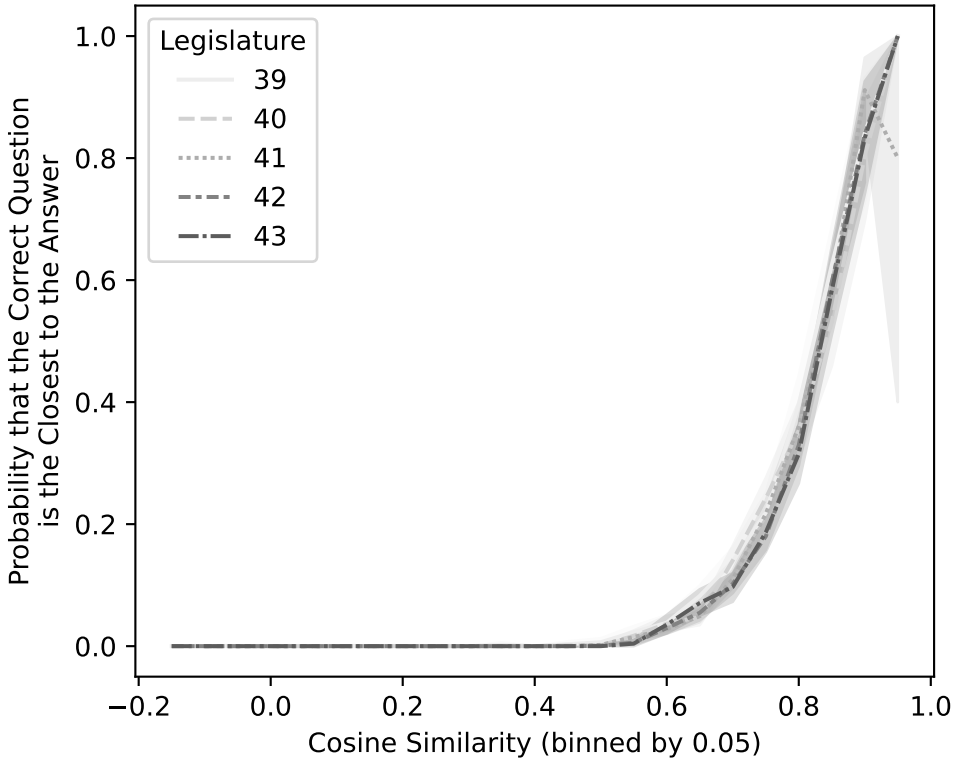


Figure A8: Probability that the Correct Question is the Closest to the Answer by Cosine Similarity Between Questions and Answers and Legislature

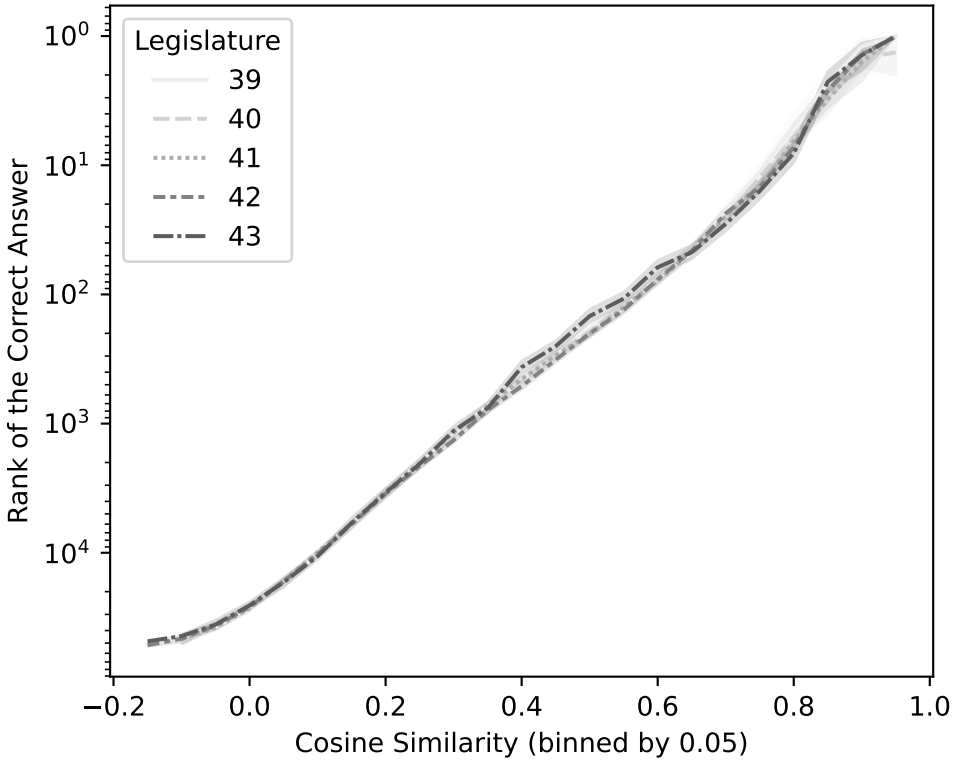


Figure A9: Rank of the Correct Answer by Cosine Similarity Between Questions and Answers and Legislature

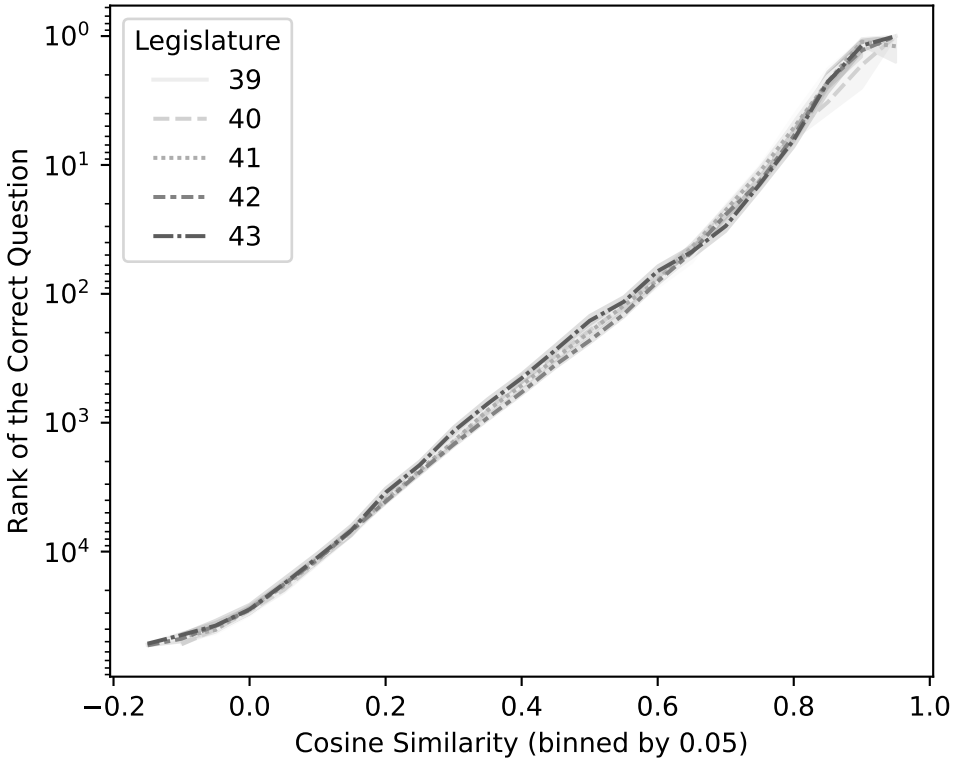


Figure A10: Rank of the Correct Question by Cosine Similarity Between Questions and Answers and Legislature

Table A5: Validity Experiment Results

Category	Average Cosine Similarity	Count
Non-Replies	0.4327	189
Intermediate Replies	0.5454	194
Full Replies	0.6268	117

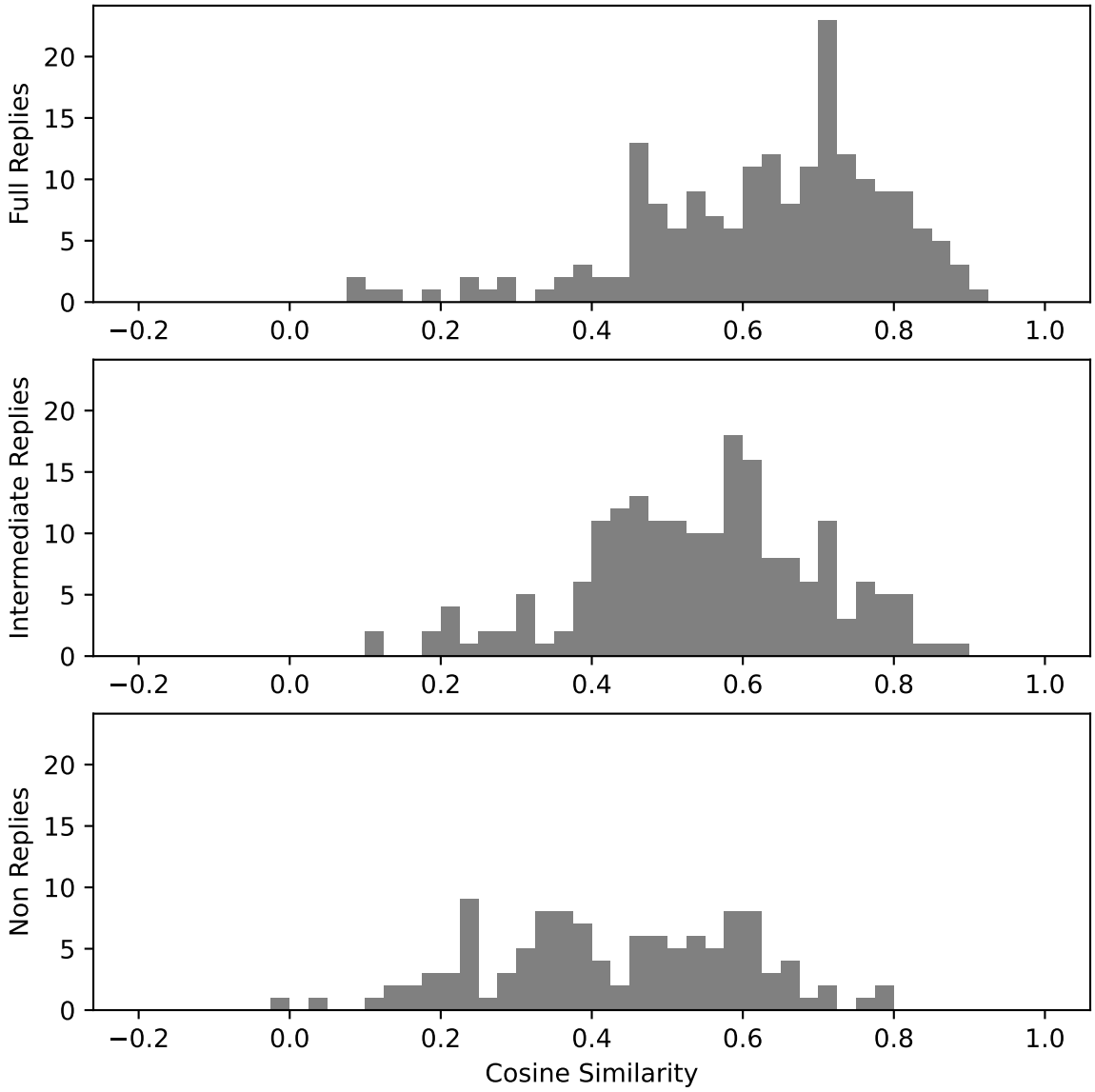
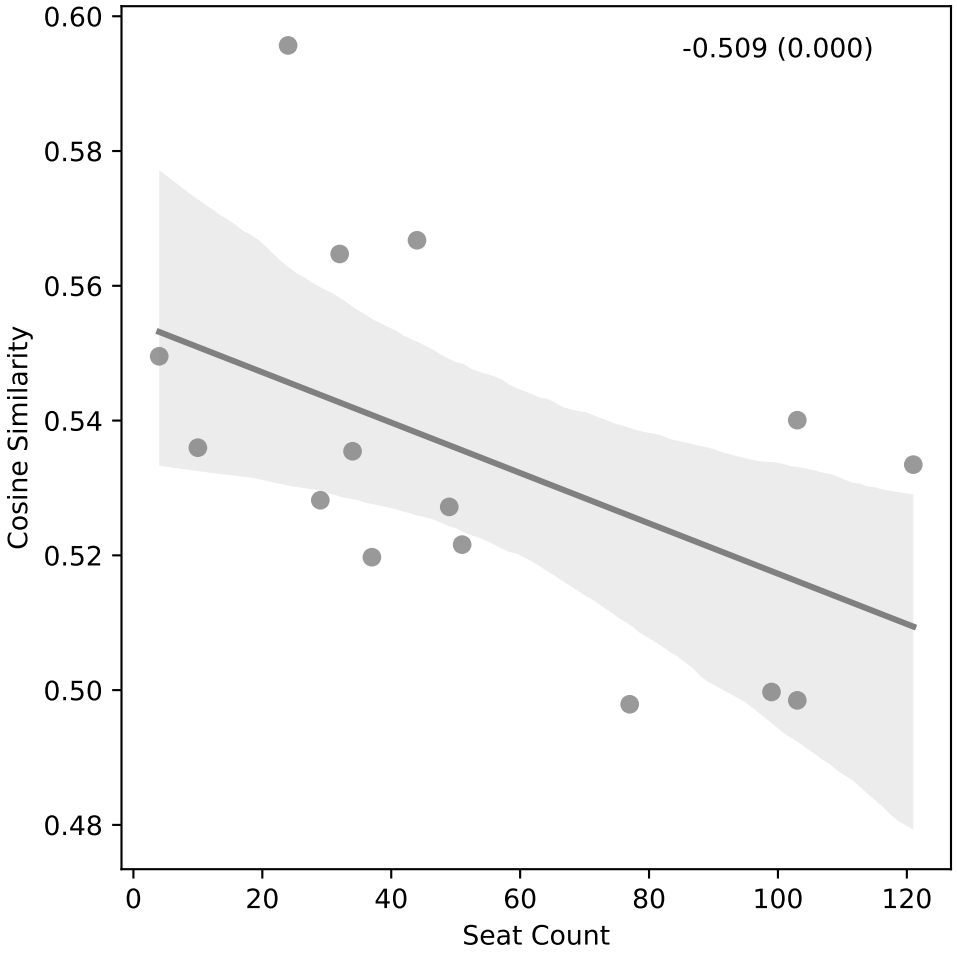


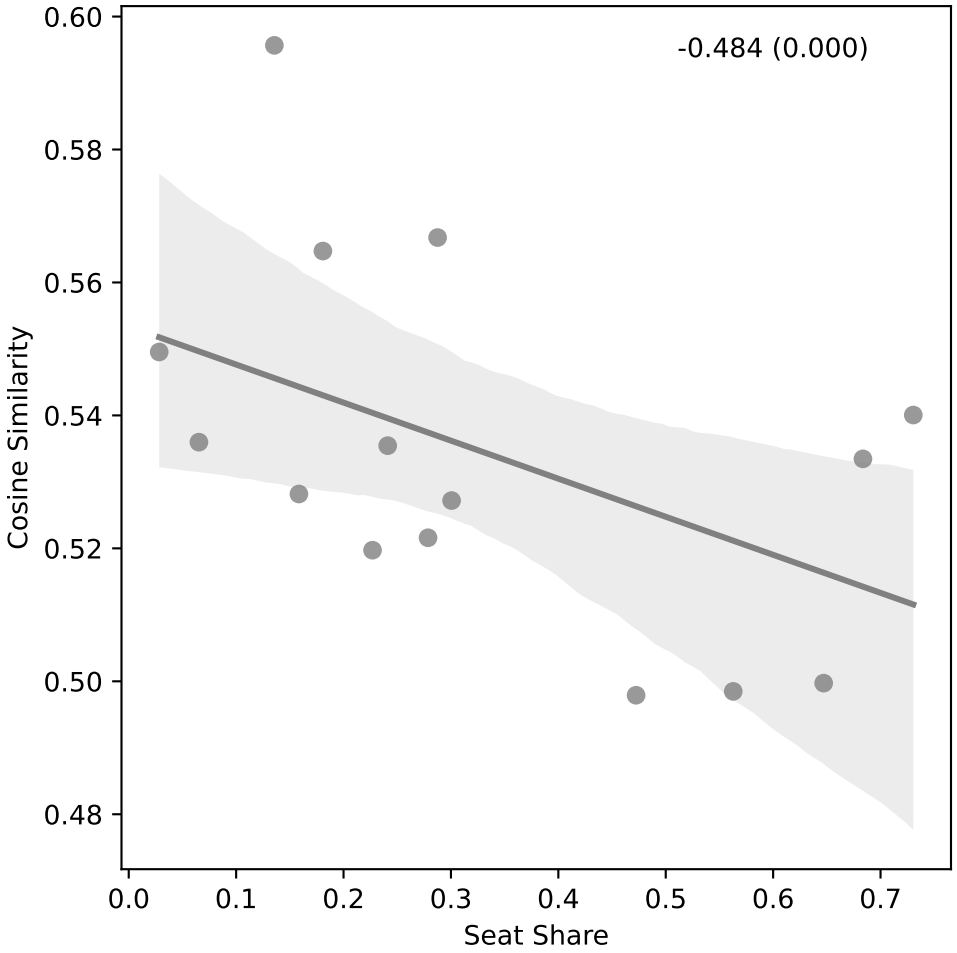
Figure A11: Distribution of the Cosine Similarity Between Questions and Answers by Reply Category



Notes:

1. This figure only includes opposition parties.
2. The seat count reflects each party's representation at the start of the legislature.
3. The correlation coefficient and corresponding p-value are shown in the top right corner.

Figure A12: Average Cosine Similarity Between Questions and Answers by Seat Count



Notes:

1. This figure only includes opposition parties.
2. The seat share reflects each party's representation at the start of the legislature.
3. The correlation coefficient and corresponding p-value are shown in the top right corner.

Figure A13: Average Cosine Similarity Between Questions and Answers by Seat Share

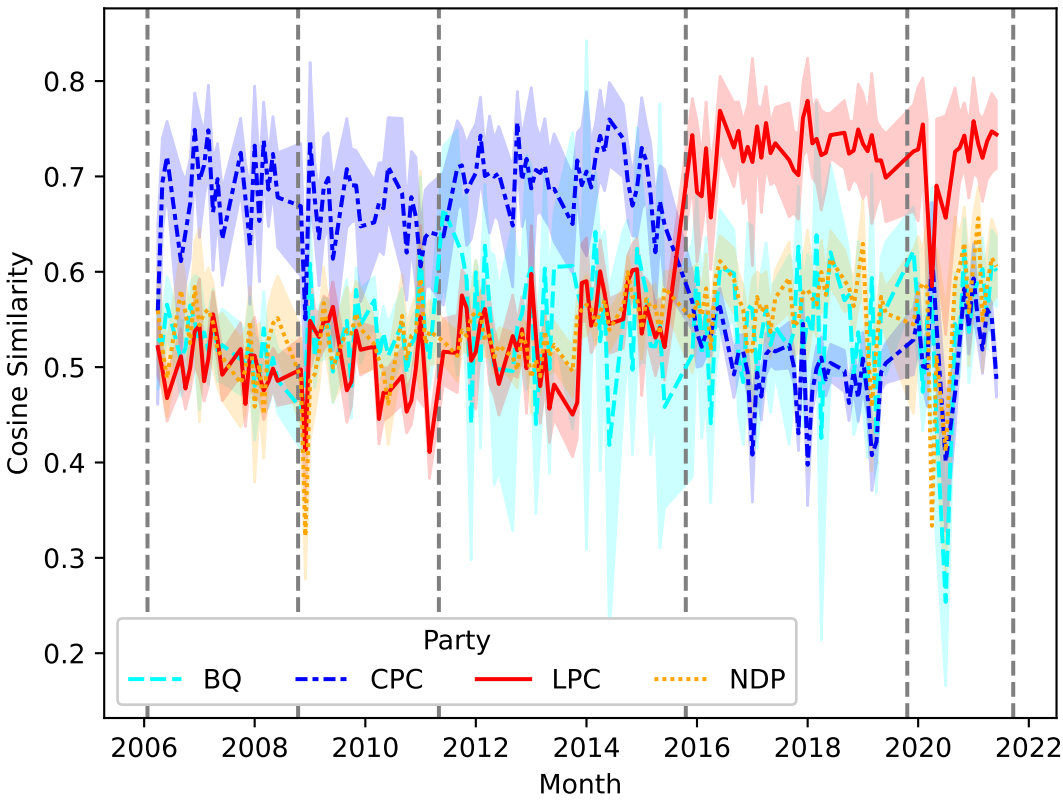


Figure A14: Monthly Evolution of the Average Cosine Similarity Between Questions and Answers by Party

System Prompt: You are a helpful, honest, and respectful assistant.

Your task is to label topics clustering questions asked by members of Parliament to Cabinet ministers during the Question Period in the Canadian House of Commons.

You must meticulously follow all the instructions you are given.

Example Prompt: I have a topic that contains the following documents:

- Traditional diets in most cultures were primarily plant-based with a little meat on top, but with the rise of industrial-style meat production and factory farming, meat has become a staple food.
- Meat, but especially beef, is the word food in terms of emissions.
- Eating meat doesn't make you a bad person, not eating meat doesn't make you a good one.

The topic is described by the following keywords: meat, beef, eat, eating, emissions, steak, food, health, processed, chicken.

Please devise a short label for this topic. I want this label to reflect the policy issue the questions are about, irrespective of their underlying sentiment.

Please capitalize this label according to standard rules for the capitalization of titles. Make sure to return only the label without additional notes.

Example Output: Environmental Impacts of Meat Consumption

Main Prompt: I have a topic that contains the following documents:

[DOCUMENTS]

The topic is represented by the following keywords: [KEYWORDS].

Please devise a short label for this topic. I want this label to reflect the policy issue the questions are about, irrespective of their underlying sentiment.

Please capitalize this label according to standard rules for the capitalization of titles. Make sure to return only the label without additional notes.

Table A6: Prompt for Generating Topic Labels

Table A7: Average Cosine Similarity Between Questions and Answers by Topic for the Conservative Party

Topic	Average Cosine Similarity	Standard Error
Mulroney-Schreiber Affair	0.3629	0.0136
Broken Promises on Income Trusts	0.3959	0.0276
Ethics and Accountability in Government	0.3993	0.0192
Government Transparency and Security Breaches	0.4087	0.0088
"Government Contracts and Fundraising Activities"	0.4123	0.0229
Cuts to Foreign Aid Funding	0.4171	0.0205
Government Advertising Spending	0.4239	0.0171
Lobbying Activities in the Canadian Government	0.4404	0.0243
Foreign Takeover of Potash Industry	0.4411	0.0179
Accountability in Government	0.4464	0.0167
PMO Scandal and Cover-Up	0.4489	0.0079
Budgetary Accountability and Transparency	0.4588	0.0205
Foreign Investment in Natural Resources	0.4621	0.0162
Accountability of Government Spending	0.4672	0.0216
Census Policy and Data Collection	0.4718	0.0212
Budget and Taxation Policy	0.4773	0.0072
Summit Spending Controversies	0.4804	0.0192
Government Accountability and Transparency	0.4837	0.0182
Opposition to Enbridge Northern Gateway Pipeline	0.4856	0.0189
Job Loss and Economic Instability	0.4858	0.0068
Election Fraud Scandal	0.4859	0.0176
Employment Insurance Reform	0.4987	0.0062
Privacy Oversight and Security Legislation	0.5059	0.0071
Quebec Securities Jurisdiction	0.5060	0.0188
Social Welfare Policy	0.5082	0.0017
Reproductive Rights and Abortion Access	0.5099	0.0155

Table A7: Average Cosine Similarity Between Questions and Answers by Topic for the Conservative Party

Topic	Average Cosine Similarity	Standard Error
Environmental Protection and Assessment Reform	0.5128	0.0169
Access to Information Reform	0.5230	0.0177
Political Interference in Cadman Affair	0.5258	0.0171
Quebec Regional Development Funding Cuts	0.5259	0.0123
Emissions Reduction Targets and Kyoto Protocol	0.5295	0.0147
Economic Growth and Job Creation Strategies	0.5297	0.0136
HST Policy in Canada	0.5302	0.0192
Forestry Industry Loan Guarantees	0.5306	0.0238
Environmental Spill Response Policy	0.5318	0.0136
Aged Workers' Support Program	0.5339	0.0254
Infrastructure Funding and Municipalities	0.5352	0.0071
Trade Protectionism and Job Losses	0.5401	0.0138
Digital Divide and Internet Access	0.5407	0.0189
Torture and Detainee Abuse in Afghanistan	0.5409	0.0055
Fuel Price Regulation	0.5409	0.0180
Climate Change Policy	0.5415	0.0050
Forestry Industry Support	0.5445	0.0146
Tax Evasion and Combating Fraud	0.5446	0.0234
Justice Appointment Process Controversy	0.5473	0.0113
Softwood Lumber Agreement	0.5475	0.0179
Election Integrity	0.5499	0.0110
Single Parent Families and Income Splitting	0.5516	0.0190
Auto Sector Crisis	0.5518	0.0210
Federal Defence Procurement	0.5531	0.0060
Electoral Reform	0.5544	0.0129
Copyright Law Reform	0.5547	0.0228

Table A7: Average Cosine Similarity Between Questions and Answers by Topic for the Conservative Party

Topic	Average Cosine Similarity	Standard Error
Drug Supply Chain Management	0.5592	0.0198
Credit Card Fee Regulation	0.5604	0.0155
Food Safety Regulation and Enforcement	0.5648	0.0103
Linguistic Rights and Governance	0.5664	0.0102
Seniors' Pension Policy	0.5724	0.0067
Labour Market Abuses and Reform	0.5788	0.0102
Housing Policy and Funding	0.5794	0.0121
Missing and Murdered Indigenous Women and Girls	0.5804	0.0122
Aboriginal Education and Reconciliation	0.5811	0.0064
Immigration Policy Reform	0.5832	0.0112
Farmers' Right to Vote on Agricultural Policy	0.5860	0.0143
Afghanistan Military Mission	0.5865	0.0064
Gender Equality Policy Attacks	0.5868	0.0108
Quebec's GST Compensation	0.5890	0.0196
Funding for Cultural Programs and Events	0.5929	0.0140
Fisheries Management and Sustainability	0.5977	0.0094
Search and Rescue Services	0.5990	0.0176
Crime Policy	0.6012	0.0071
Northern Food Security	0.6013	0.0185
Child Care Policy	0.6027	0.0105
Medical Isotope Supply Crisis	0.6027	0.0128
Rail Safety and Enforcement	0.6047	0.0098
Rail Service and Safety	0.6058	0.0249
Supply Management Policy	0.6063	0.0150
Research Funding Cuts in Canada	0.6066	0.0183
Agricultural Policy and Farmers' Interests	0.6096	0.0106

Table A7: Average Cosine Similarity Between Questions and Answers by Topic for the Conservative Party

Topic	Average Cosine Similarity	Standard Error
Youth Unemployment and Job Creation	0.6104	0.0145
Repatriation of Omar Khadr	0.6145	0.0120
CBC Funding and Support	0.6176	0.0117
Refugee Policy and Services	0.6260	0.0107
Health Care Wait Times	0.6261	0.0212
Trade Policy	0.6303	0.0119
Aid to Africa	0.6368	0.0152
Home Delivery Service Cuts	0.6413	0.0102
Gun Control Policy	0.6415	0.0095
Veterans' Services and Support	0.6423	0.0049
Health Care Accord Renewal	0.6437	0.0135
Bridge Tolls and Infrastructure Funding	0.6481	0.0132
Vaccine Distribution and Availability	0.6506	0.0105
Military Sexual Misconduct	0.6683	0.0202
International Threats and Security	0.6726	0.0161
Ukraine Policy and Support	0.6913	0.0186
Humanitarian Aid and Conflict Resolution	0.7066	0.0081

Table A8: Average Cosine Similarity Between Questions and Answers by Topic for the Liberal Party

Topic	Average Cosine Similarity	Standard Error
Ethics Violations and Financial Accountability	0.3649	0.0161
Financial Conflict of Interest	0.3906	0.0245
Attorney General Scandal	0.4126	0.0068
National Security Review of Foreign Takeovers	0.4208	0.0172
Ethics and Access to Government	0.4353	0.0090
Criminal Justice Reform	0.4381	0.0230
Jobs Crisis in Canada	0.4383	0.0168
Government Accountability on India Trip Controversy	0.4408	0.0168
Political Interference in the Mark Norman Case	0.4503	0.0131
Huawei Ban and Security Concerns	0.4509	0.0234
Defence Minister's Credibility Crisis	0.4600	0.0181
Apologetics and Respect for Official Languages	0.4615	0.0249
Cronyism in Fisheries Allocation	0.4644	0.0234
Financial Conflicts of Interest	0.4650	0.0105
National Security and Scientific Collaboration	0.4661	0.0165
Rail Blockades and Economic Impact	0.4674	0.0182
Government Expense Accountability	0.4731	0.0247
Federal-Provincial Relations in Quebec	0.4737	0.0180
Carbon Tax Policy	0.4757	0.0071
Pipeline Policy and Job Losses	0.4871	0.0184
Budget and Fiscal Responsibility	0.4876	0.0051
Energy Policy and Pipeline Development	0.4997	0.0063
Taxation of Digital Giants	0.5086	0.0241
Terrorism Reintegration and Security Policy	0.5128	0.0152
Public Sector Pay System Crisis	0.5133	0.0154
Accountability and Transparency in Government	0.5279	0.0025

Table A8: Average Cosine Similarity Between Questions and Answers by Topic for the Liberal Party

Topic	Average Cosine Similarity	Standard Error
Privacy and Data Protection	0.5303	0.0112
Referendum on Electoral Reform	0.5312	0.0085
Employment Insurance Reform	0.5322	0.0248
Immigration Policy and Border Control	0.5337	0.0084
Retirement Security Reform	0.5343	0.0162
Indigenous Children's Rights	0.5360	0.0123
Infrastructure Privatization	0.5381	0.0111
Disability Tax Credit Access Denials	0.5416	0.0189
Quarantine Policy and Enforcement	0.5466	0.0108
Softwood Lumber Industry Policy	0.5466	0.0129
Freedom of Expression Under Attack	0.5528	0.0127
#MeToo in the Canadian Military	0.5556	0.0106
Rapid COVID-19 Testing Availability	0.5561	0.0139
Canada-China Relations	0.5564	0.0151
Parole System Reform	0.5568	0.0149
Climate Change Policy	0.5663	0.0112
Medical Assistance in Dying Legal Challenges	0.5706	0.0165
Health Care Funding	0.5776	0.0121
Tax Evasion and Conflict of Interest	0.5795	0.0101
Fighter Jet Replacement Policy	0.5810	0.0136
Indigenous Rights and Consultation	0.5853	0.0133
Combat Mission Deployment	0.5866	0.0095
Indigenous Financial Transparency	0.5880	0.0195
Human Rights and Arms Exports to Saudi Arabia	0.5886	0.0167
Trade Agreement Negotiations and Job Losses	0.5895	0.0142
Trade Agreements and Job Protection	0.6037	0.0153

Table A8: Average Cosine Similarity Between Questions and Answers by Topic for the Liberal Party

Topic	Average Cosine Similarity	Standard Error
Agricultural Policy and Supply Management	0.6062	0.0074
Marijuana Legalization and Criminal Records	0.6109	0.0136
Middle East Policy and Human Rights	0.6140	0.0096
Appointment Process and Ethics	0.6159	0.0103
Shipbuilding Contract Dispute	0.6160	0.0175
Trade Tariffs and Job Losses	0.6192	0.0175
Child Care Policy	0.6194	0.0136
Small Business Support	0.6293	0.0129
Veteran Benefits and Entitlements	0.6301	0.0094
Aerospace Industry Support	0.6329	0.0075
Indigenous Drinking Water Advisories	0.6360	0.0159
Vaccine Supply Delays and Canada's Ranking	0.6365	0.0064
Indigenous Women's Inquiry	0.6384	0.0142
Job Loss and Economic Impact in Oshawa	0.6392	0.0204
Aging Population and Social Security	0.6397	0.0137
Rural Broadband Access	0.6424	0.0135
Refugee Resettlement Policy	0.6465	0.0213
Opioid Epidemic	0.6476	0.0186
Housing Policy and Investments	0.6506	0.0091
Salmon Conservation and Management	0.6573	0.0119
Public Safety and Gun Control	0.6624	0.0136
Federal Language Policy in Quebec	0.6626	0.0096
Pharmacare Policy	0.6640	0.0150
Gender-Based Violence and Empowerment	0.6657	0.0110
Rail Safety	0.6829	0.0130
Mental Health Services and Accessibility	0.7066	0.0142

B Issue Ownership

To appreciate the relative reputation of Canadian political parties on policy issues, we consider responses to the question "Which party would do the best job at handling each of the following issues?" in the last three editions of the Canadian Election Study (Stephenson et al. 2020, 2022). The policy issues considered in all three editions are the following: Crime and Justice, Defense, Education, Environment, Healthcare, Immigration and Minorities, and International Diplomacy. The distribution of responses is depicted in Figure A15. The relative status of parties over policy issues varies over time. Nevertheless, between the Conservative Party and the Liberal Party, the former has consistently enjoyed a relatively better reputation for Crime and Justice and Defense. In contrast, the latter has a better reputation for Education, Environment, Healthcare, Immigration and Minorities, and International Diplomacy.

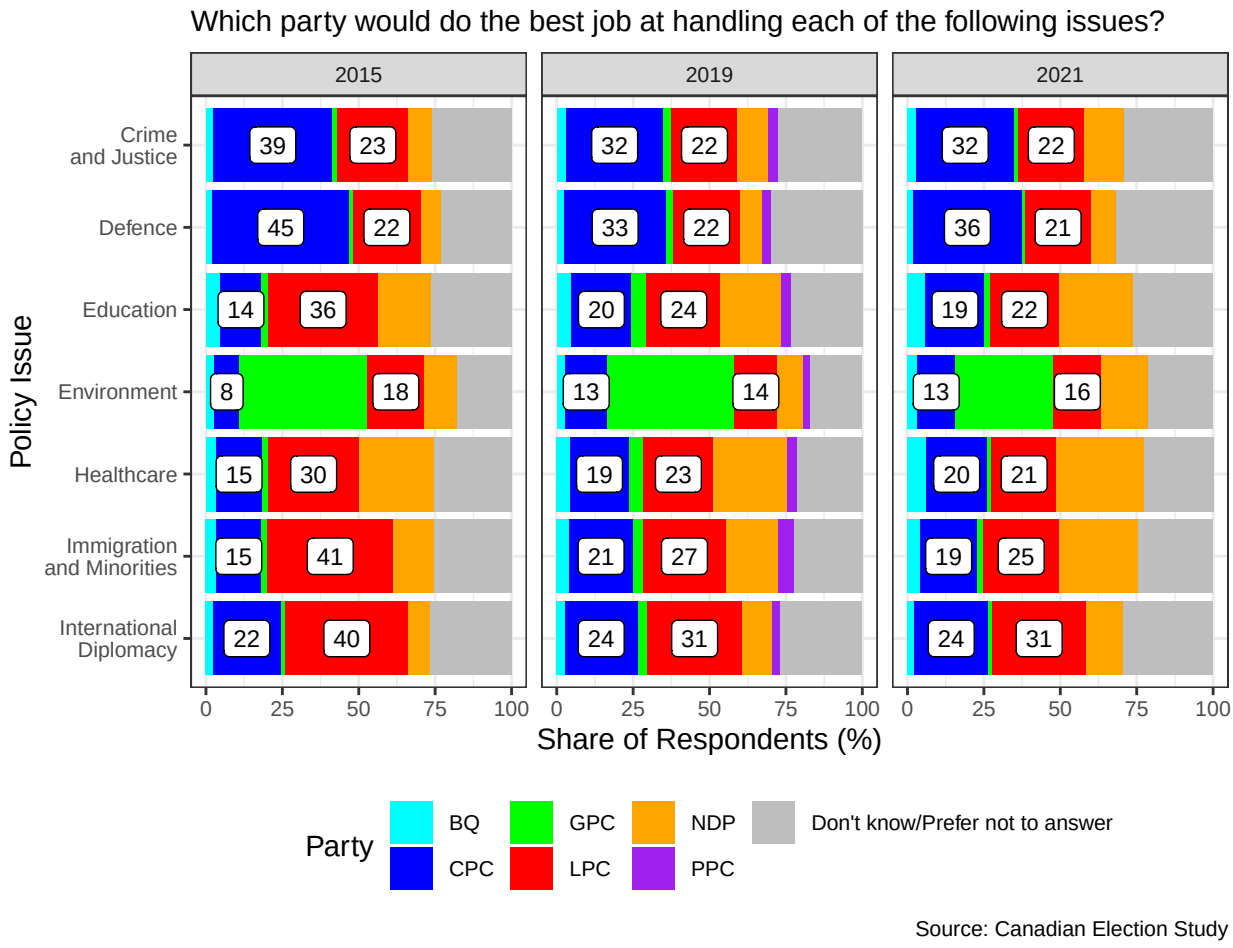


Figure A15: Reputation of Political Parties Over Policy Issues

C Robustness Check: Pre-Trained Model

In this section, we present results from the pre-trained model without fine-tuning.

Figure A16 depicts the distribution of cosine similarity estimates between questions and their answers. For reference, the null distribution of the cosine similarity between questions and random answers is shown using hollow lines. Descriptive statistics for this distribution are provided in Table A9. Remarkably, the distribution of cosine similarity estimates for the pre-trained model is much closer to the corresponding null distribution than the cosine similarity distribution for the fine-tuned model. In other words, the pre-trained model ascribes a much lower relative relevance to observed answers than the fine-tuned model. To a meaningful extent, this suggests using the fine-tuned model for our analysis. Additionally, Figure A17 compares cosine similarity estimates between the pre-trained and fine-tuned models, showing they have a significant but moderate correlation.

Figure A18 illustrates the average cosine similarity between questions and answers conditional on the legislature and the party affiliation of the MP posing the question. The observed patterns closely mirror those from the fine-tuned model, implying that these substantive findings are resilient to using the pre-trained model without fine-tuning.

Figure A19 illustrates the average cosine similarity between questions and answers, conditioned on the party affiliation and portfolio of the government member responding. Additionally, Figures A20 and A21 present the 20 topics of questions with the lowest and highest average cosine similarity between questions and answers for the Conservative and Liberal parties, respectively. The observed patterns are generally consistent with those from the fine-tuned model. Yet statistically significant differences in party-neutral average cosine similarities are observed with the pre-trained model, but not the fine-tuned model, for the following portfolios: Democratic Institutions, Finance, Fisheries and Ocean, International Development, and Veterans Affairs. In contrast, statistically significant differences are found with the fine-tuned model, but not the pre-trained model, for the following portfolios: Agriculture and Agri-Food, Environment, Foreign Affairs, and Justice. Topic-level results involve fewer changes, except in the relative ordering of topics.

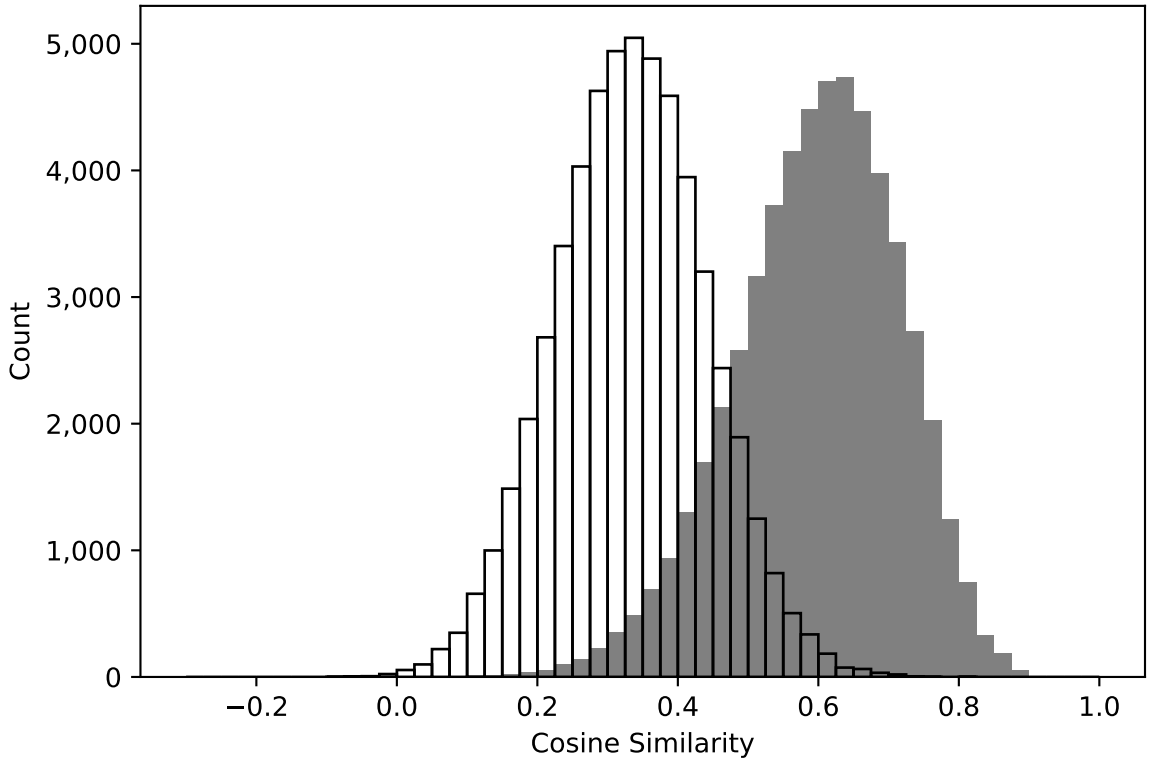


Figure A16: Distribution of the Cosine Similarity Between Questions and Answers

Table A9: Descriptive Statistics of the Distribution of the Cosine Similarity Between Questions and Answers

Count	54,914
Mean	0.5975
Standard Deviation	0.1163
Minimum	0.0923
First Quartile	0.5234
Median	0.6061
Third Quartile	0.6807
Maximum	0.9274

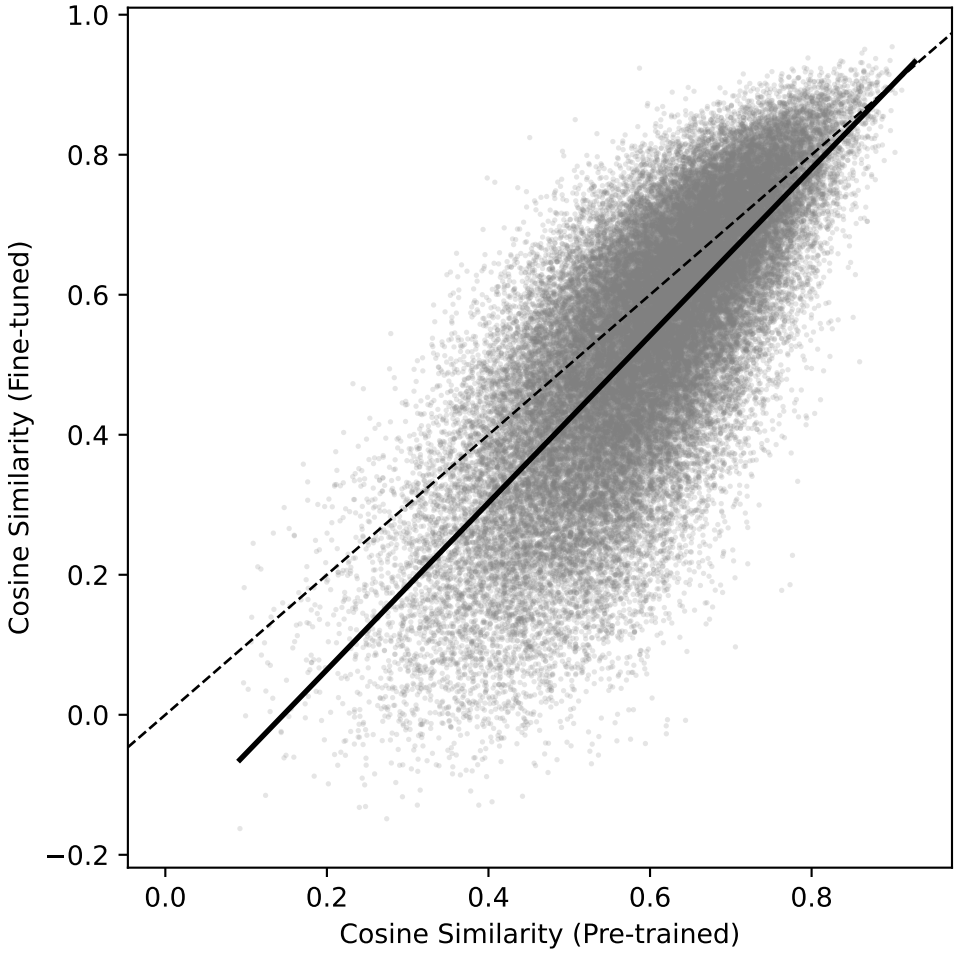


Figure A17: Comparison of Cosine Similarity Estimates Between the Pre-Trained and Fine-Tuned Models

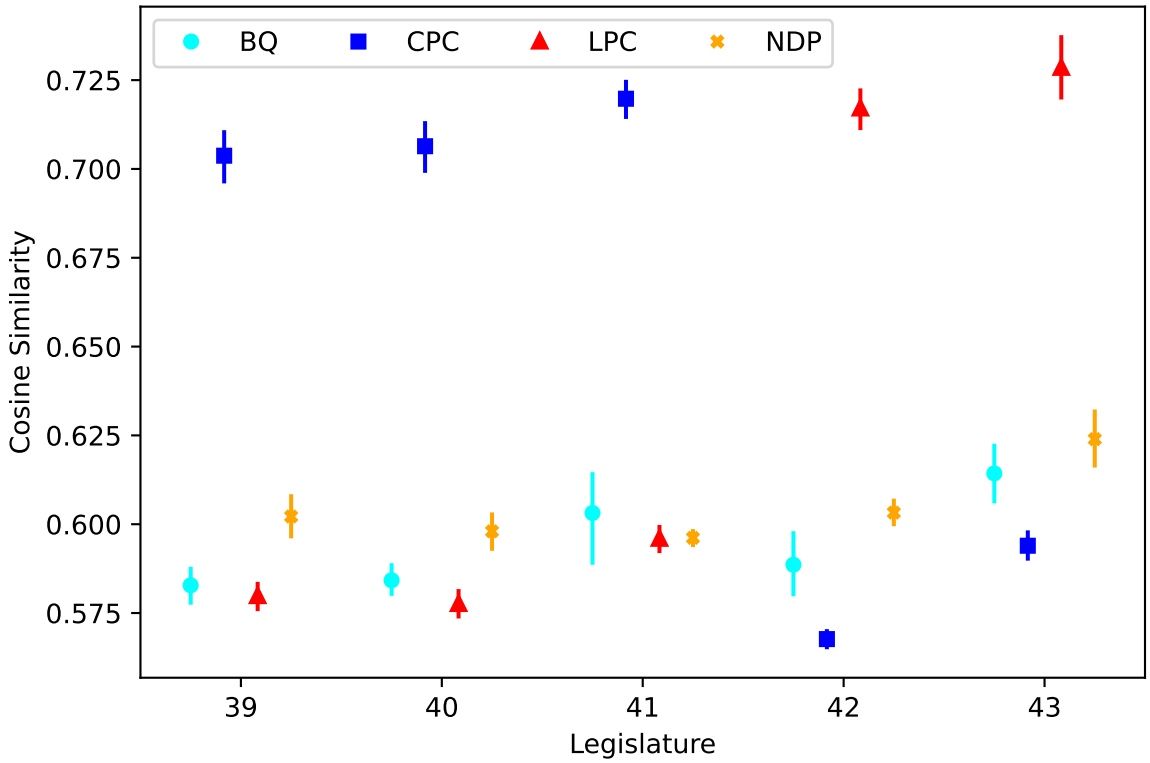


Figure A18: Average Cosine Similarity Between Questions and Answers by Party and Legislature

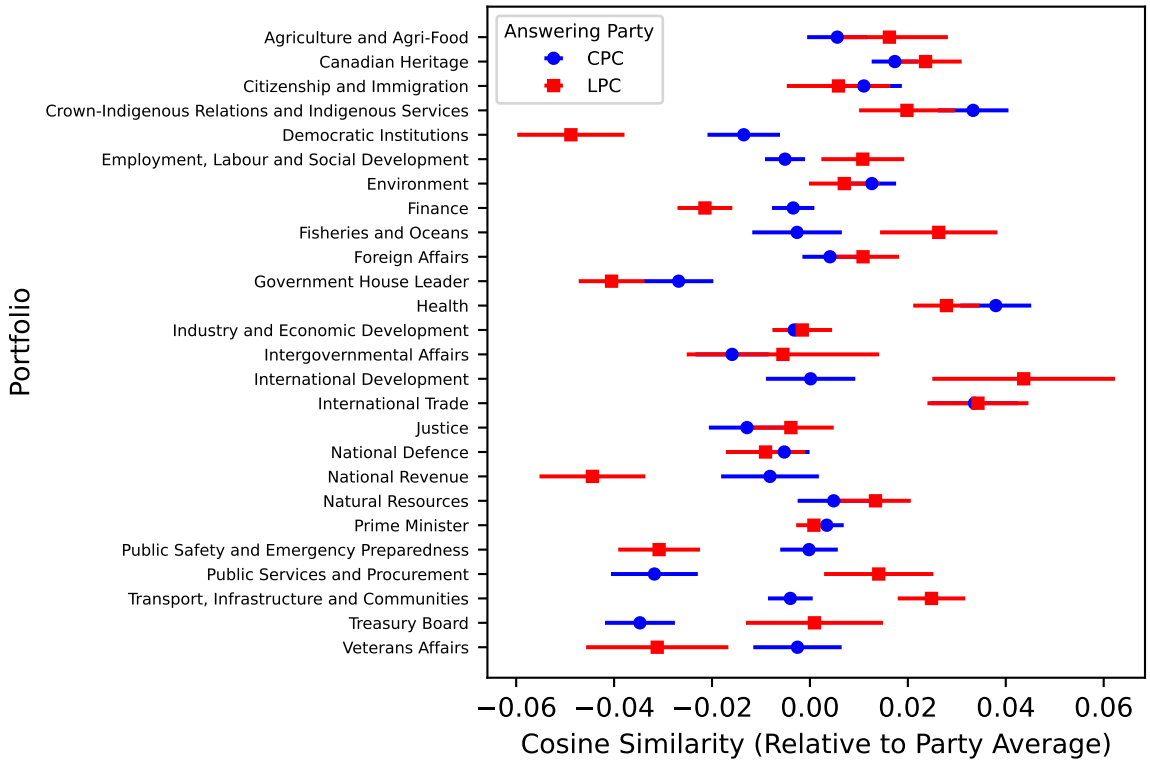


Figure A19: Average Cosine Similarity Between Questions and Answers by Party and Portfolio

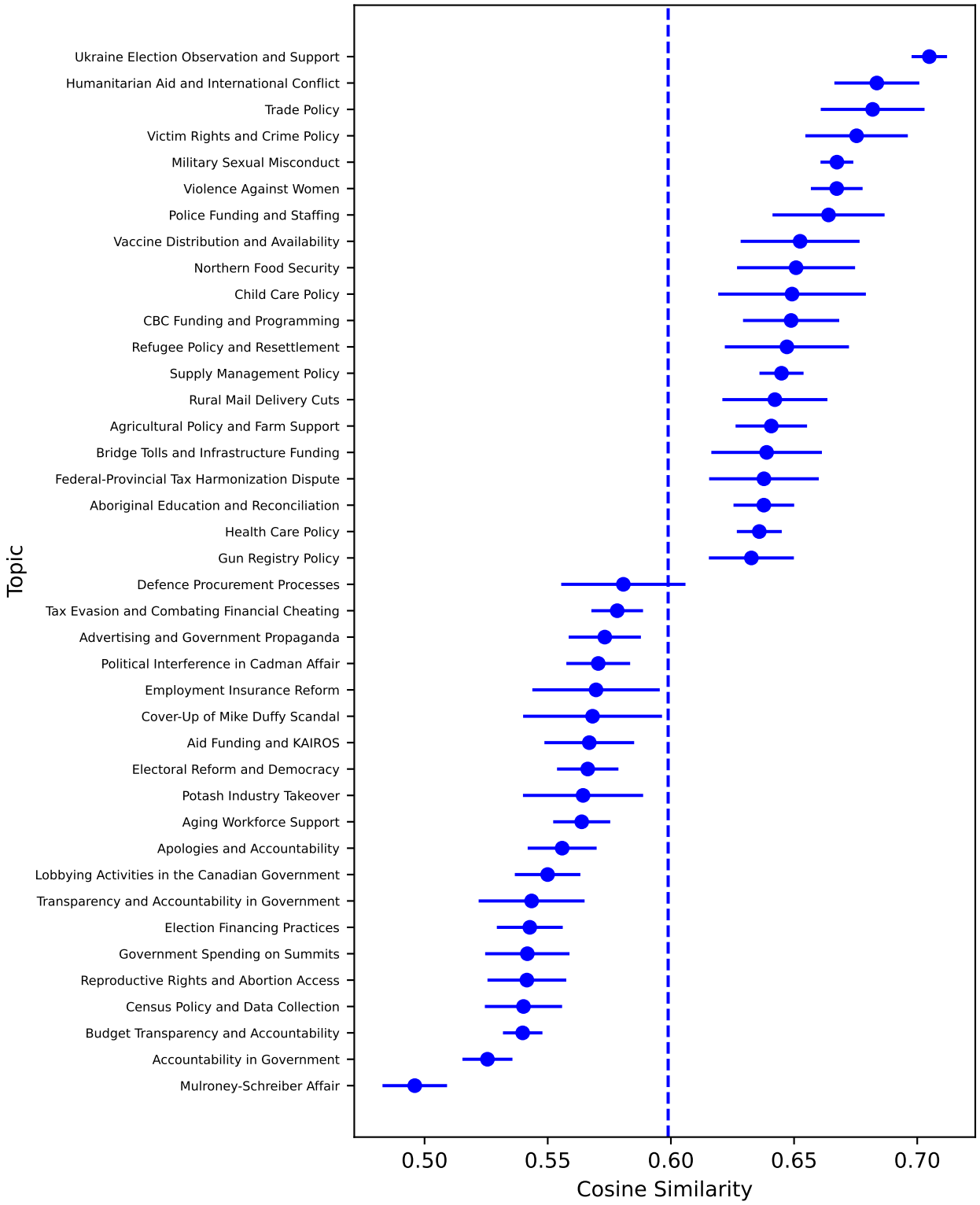


Figure A20: Topics with the 20 Lowest and Highest Average Cosine Similarity Between Questions and Answers for the Conservative Party

Figure A21: Topics with the 20 Lowest and Highest Average Cosine Similarity Between Questions and Answers for the Liberal Party

D Robustness Check: Document Length

A potential issue with using distance metrics between estimated latent representations as a measurement is that sampling errors mechanically increase their distance and lower their similarity. While this issue affects all latent representations and distance metrics, it is particularly pronounced when dealing with high-dimensional representations, as we do in this article. This has been carefully explored and documented in previous literature (Gentzkow, Shapiro, and Taddy 2019; Loon et al. 2022; Green et al. 2024).

This issue may result in a correlation between the cosine similarity and the lengths of questions and answers. The intuition behind this is that shorter questions or answers have larger sampling errors in their latent representations due to being estimated with less information. As a result, the cosine similarity between shorter questions and answers is likely to exhibit a downward bias. If there are systematic differences in the lengths of questions and answers across different parties and legislatures, this bias could distort our substantive findings.

Figures A22 and A23 illustrate the average cosine similarity as a function of question and answer lengths. They confirm a statistically significant correlation between the cosine similarity and question and answer lengths. This relationship is downward-sloping for question length, meaning that lengthier questions tend to have a lower cosine similarity. This contradicts what we would expect if sampling error introduced a significant bias in cosine similarity estimates. Conversely, the relationship is upward-sloping for answer length, suggesting that either longer answers have a lower sampling error, longer answers are more relevant to the initial questions, or both.

The potential downward bias in cosine similarity could affect our substantive findings regarding the relationship between answer quality and the party affiliation of the member of Parliament asking the question, but only if there are systematic differences in question and answer lengths across the latter. Figures A24 and A25 reveal systematic variations in the lengths of questions and answers based on the party affiliation of the member of Parliament asking the question and the legislature. Furthermore, Figures A26 and A27 indicate an apparent relationship between estimates of the average cosine similarity and the lengths of questions and answers, depending on the party affiliation of the member of Parliament asking the question and the legislature. This suggests that our substantive findings might be driven, at least partly, by systematic differences in the lengths of questions and answers. This could be symptomatic of a downward bias in cosine similarity resulting from sampling error.

To mitigate and evaluate the robustness of our substantive findings to any systematic relationship between cosine similarity and the lengths of questions and answers, we calculate the average cosine similarity between questions and answers, conditional on the party affiliation of the member of Parliament asking the question and the legislature, after controlling for question and answer lengths. Formally, adjusted average cosine similarity estimates are computed from a linear regression model that includes question and answer lengths and party-legislature fixed effects as covariates. Predictions are calculated for our inference dataset's average question and answer lengths. Thus, they reflect the average cosine similarity if question and answer lengths were the same across all these groups.

Figure A28 shows the estimated average cosine similarity between questions and answers, broken down by party and legislature, after controlling for both question and answer lengths. Figure A29 shows the average cosine similarity between questions and answers, categorized by the party and portfolio of the responding government member, after controlling for both question and answer lengths. The patterns mirror those described in the main text, except that Figure A29 has larger standard errors. This confirms that our main conclusions are robust and not influenced by systematic variations in question and answer lengths.

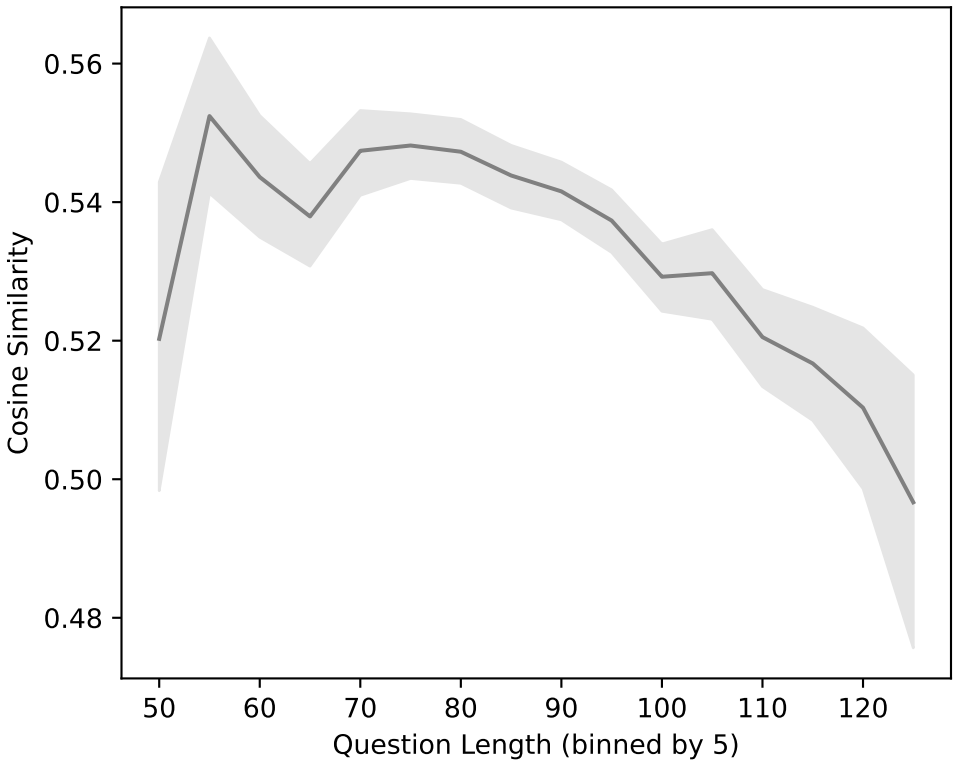


Figure A22: Average Cosine Similarity Between Questions and Answers by Question Length

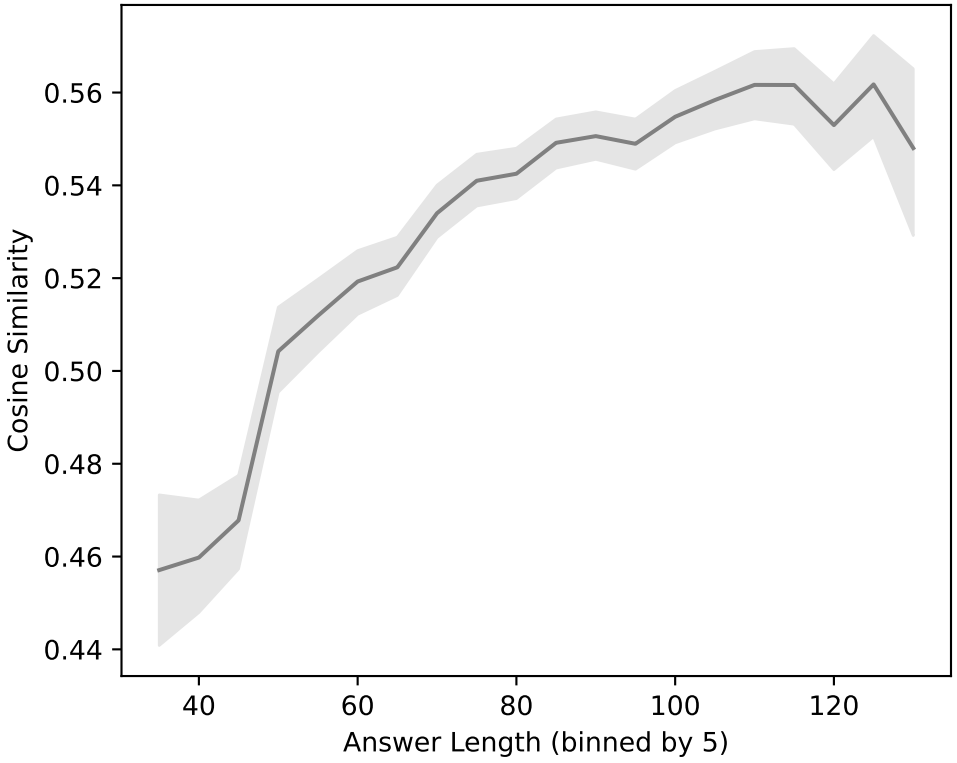


Figure A23: Average Cosine Similarity Between Questions and Answers by Answer Length

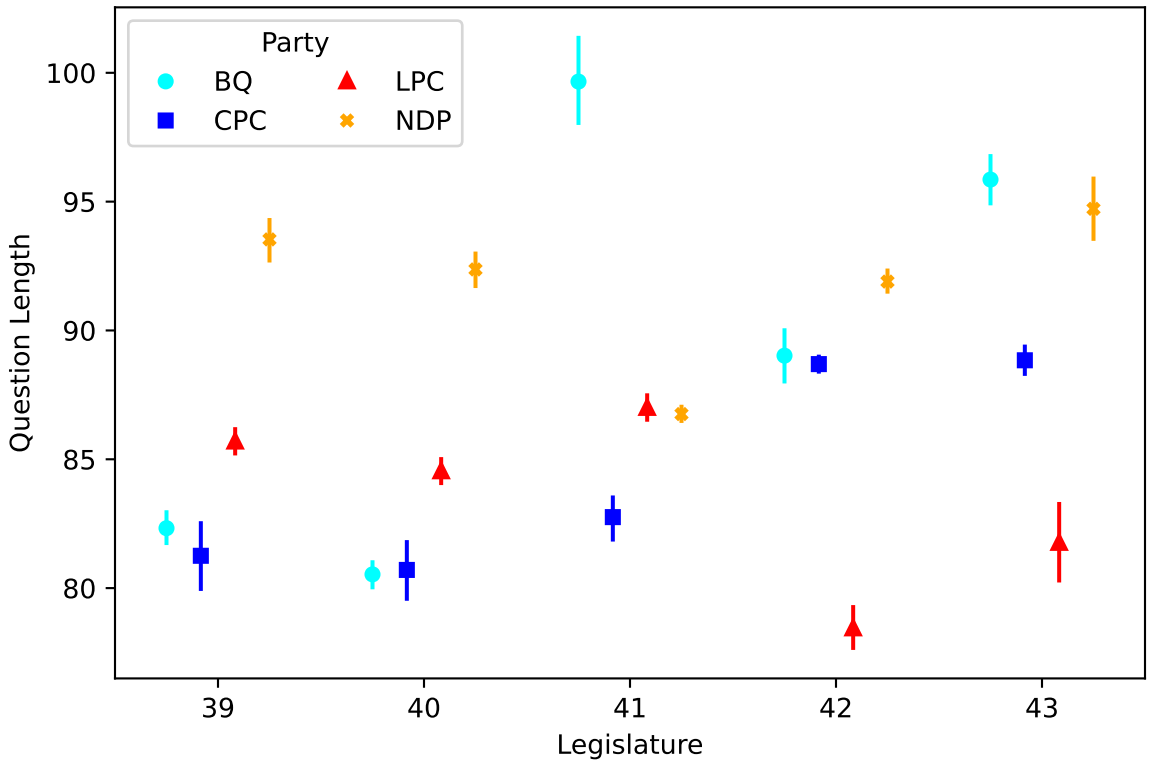


Figure A24: Average Question Length by Party and Legislature

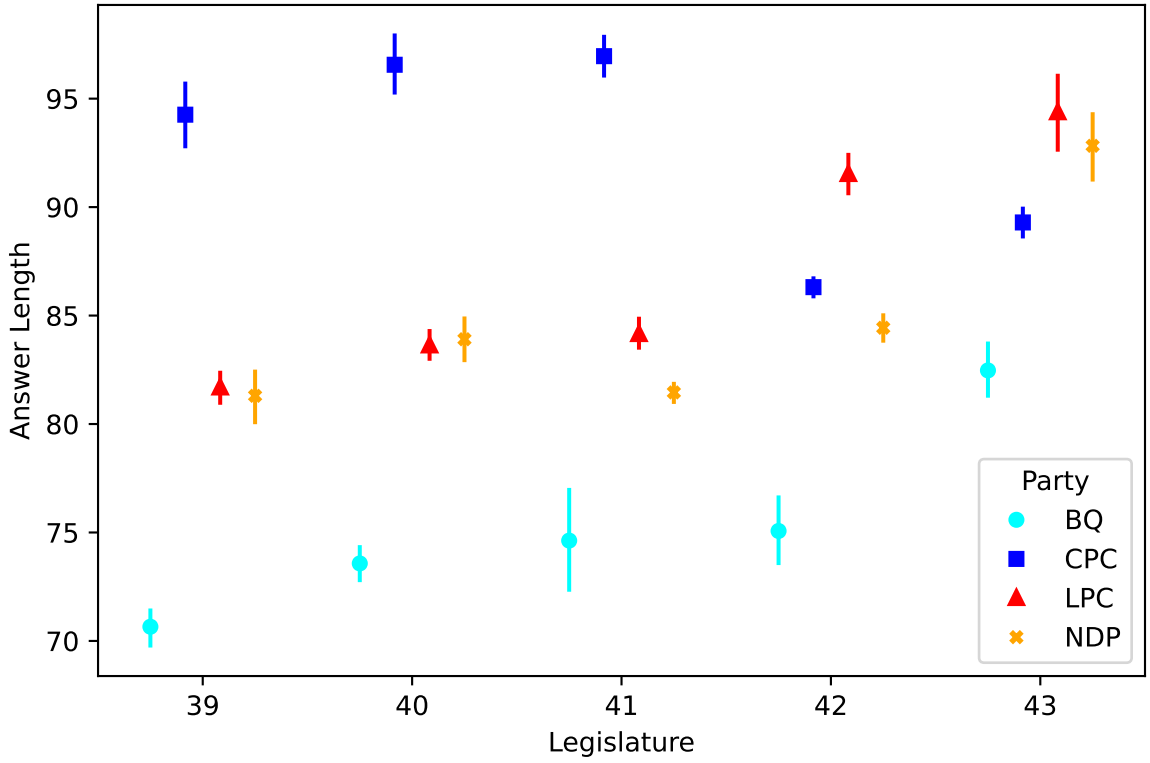


Figure A25: Average Answer Length by Party and Legislature

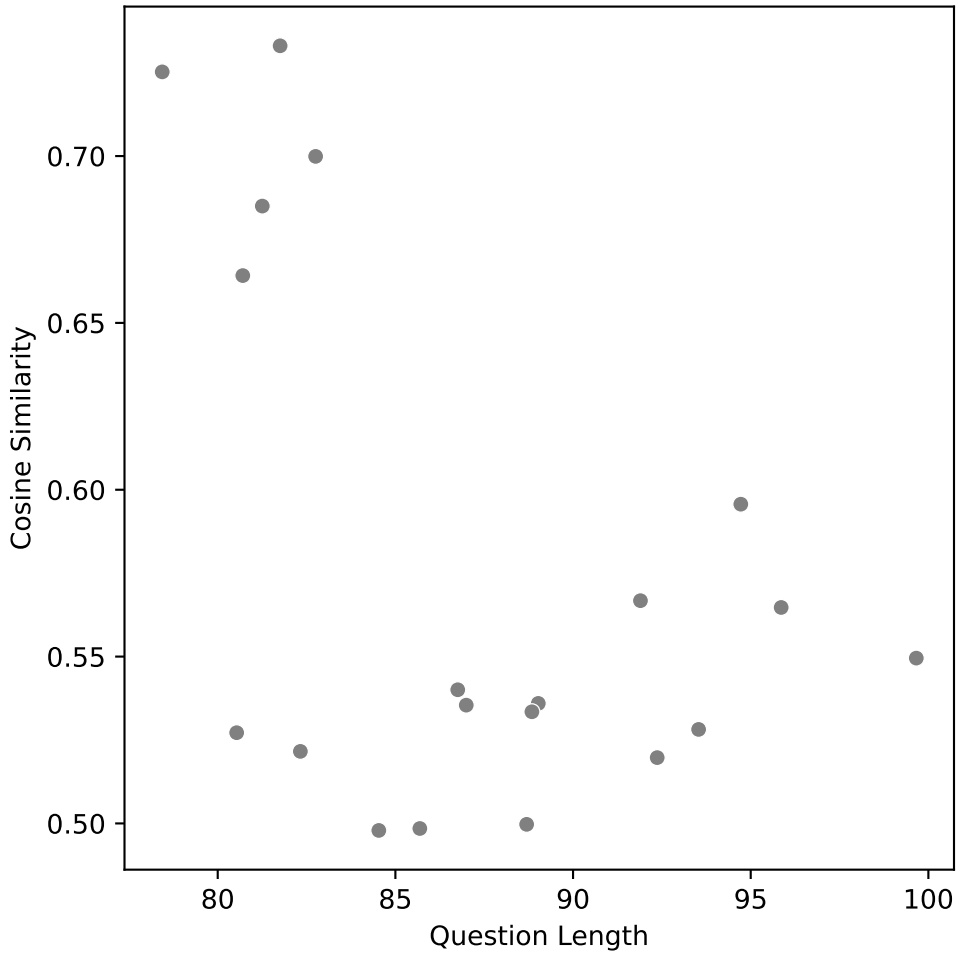


Figure A26: Average Cosine Similarity Between Questions and Answers by Average Question Length

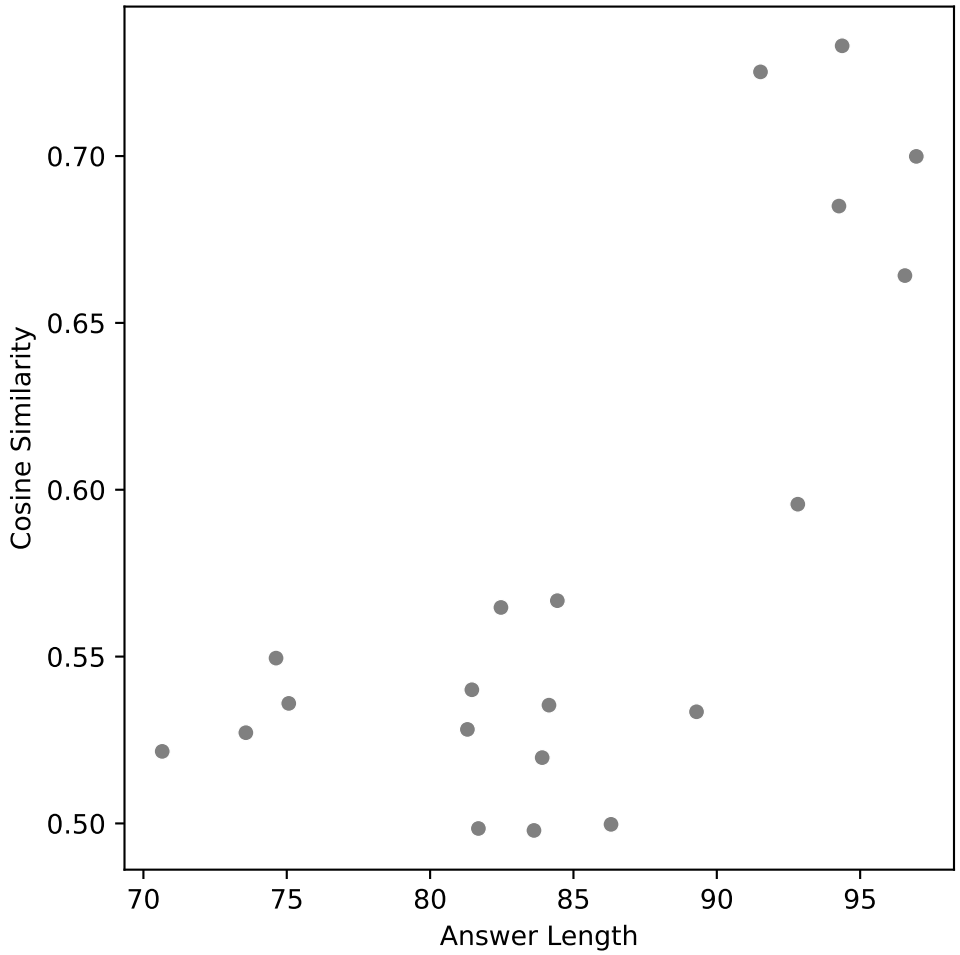


Figure A27: Average Cosine Similarity Between Questions and Answers by Average Answer Length

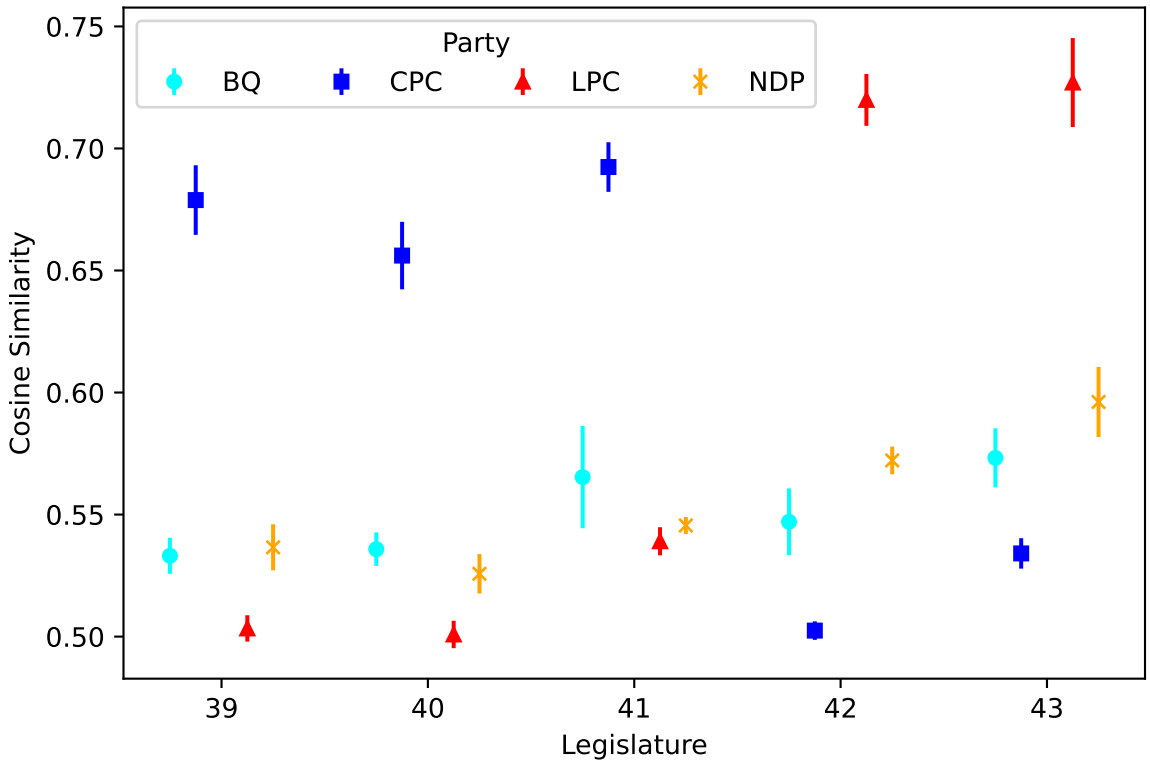


Figure A28: Average Cosine Similarity Between Questions and Answers by Party and Legislature (After Controlling for Question and Answer Lengths)

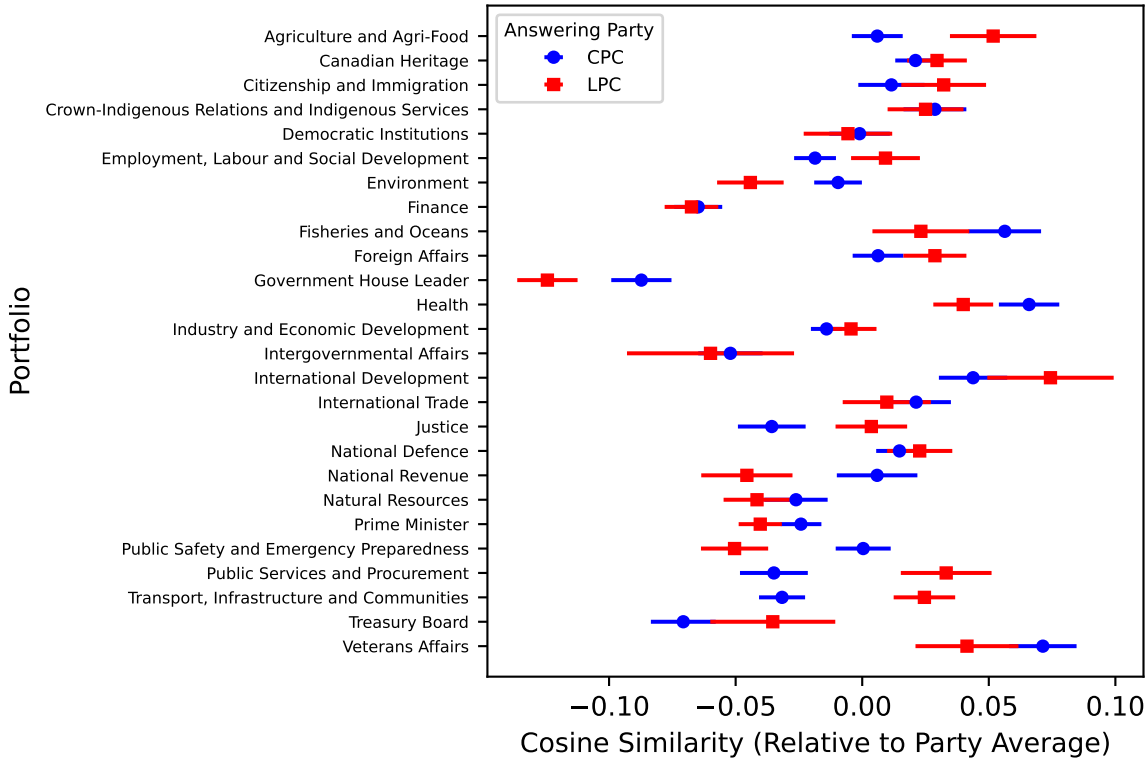


Figure A29: Average Cosine Similarity Between Questions and Answers by Party and Portfolio (After Controlling for Question and Answer Lengths)

E Robustness Check: Backbench Government MPs

To assess the impact of the inclusion of exchanges instigated by backbench government MPs in our training set on our substantive findings, we perform a data ablation study. In this analysis, we fine-tune our model on a training set that excludes questions from these members while keeping all other training hyperparameters identical to those used to train our baseline model.

Figure A30 illustrates the distribution of cosine similarity estimates generated by the model trained without questions from backbench government MPs in the training data. This distribution also excludes exchanges prompted by questions from backbench government MPs. Descriptive statistics of this distribution are listed in Table A10. Figure A31 compares the cosine similarity estimates from the main model with those from the model trained without questions from backbench government MPs. The estimates are strongly correlated, with a coefficient of 0.9121.

Figure A32 displays the average cosine similarity by the legislature and the party affiliation of the questioning MP. Figure A33 shows the average cosine similarity conditional on the party affiliation and the portfolio of the Cabinet minister or parliamentary secretary answering the question. Figures A34 and A35 present the 20 topics of questions with the lowest and highest average cosine similarity between questions and answers for the Conservative and Liberal parties, respectively. These figures show that our core substantive findings remain robust, even when exchanges involving government backbench MPs are excluded from the training set. Notably, our observation that the Conservative Party tends to address controversial issues over which it lacks a clear reputational advantage continues to hold.

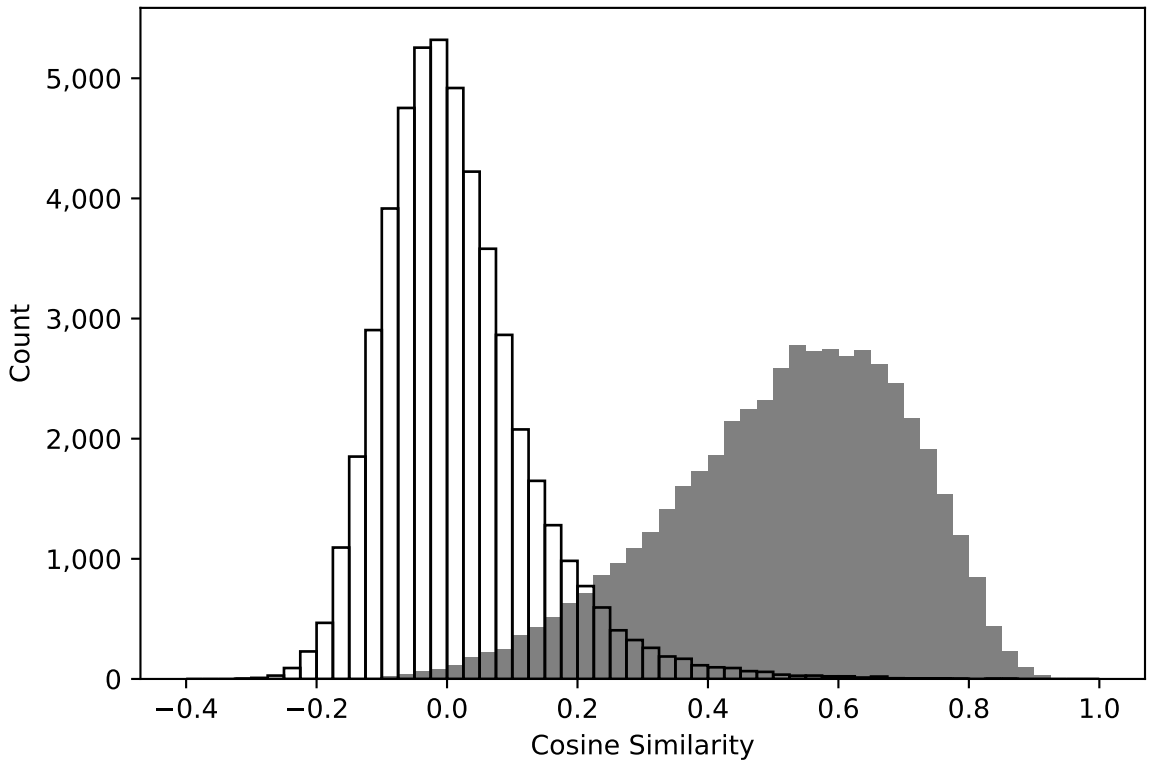


Figure A30: Distribution of the Cosine Similarity Between Questions and Answers

Table A10: Descriptive Statistics of the Distribution of the Cosine Similarity Between Questions and Answers

Count	50,818
Mean	0.5298
Standard Deviation	0.1814
Minimum	−0.1625
First Quartile	0.4146
Median	0.5521
Third Quartile	0.6665
Maximum	0.9382

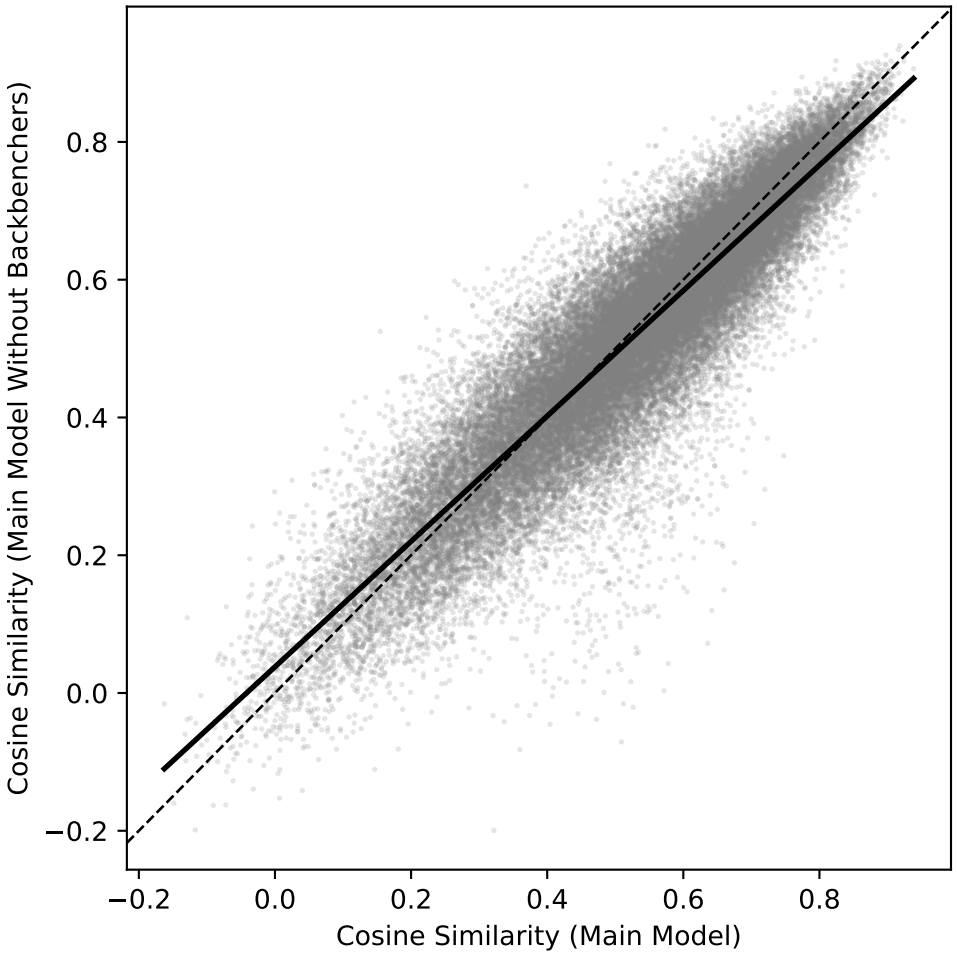


Figure A31: Comparison of Cosine Similarity Estimates Between the Main Model and Model Without Government Backbenchers

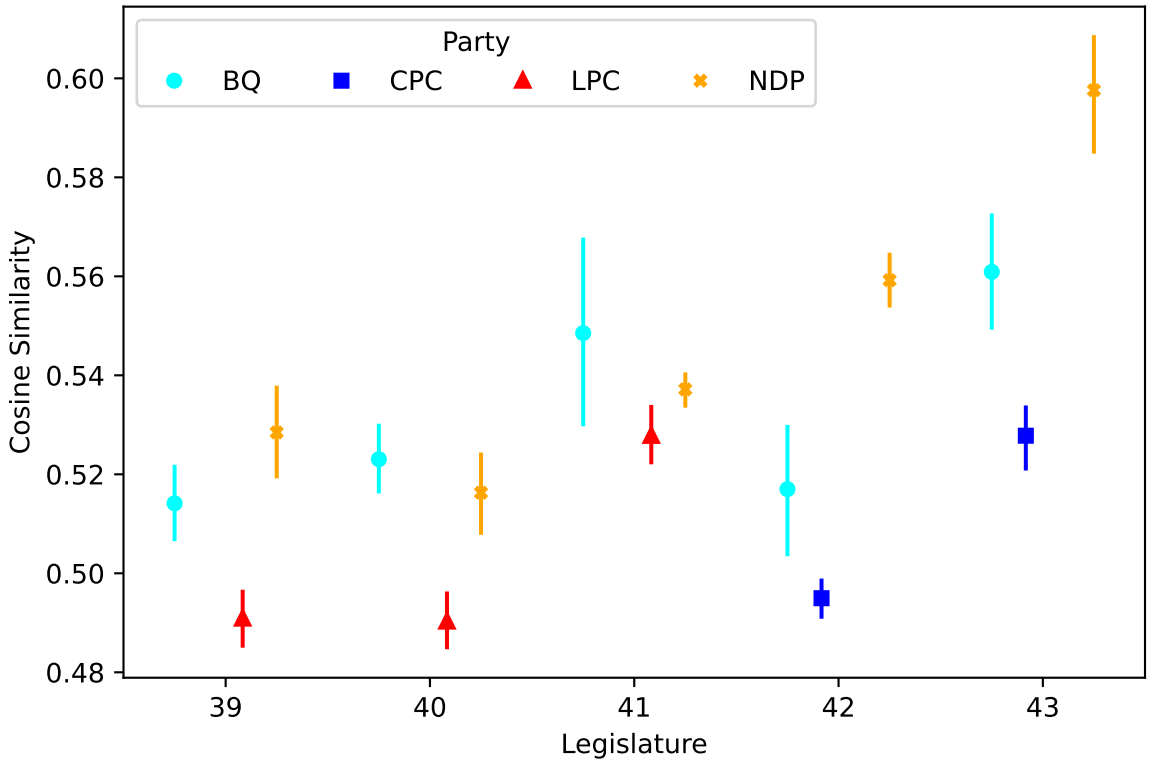


Figure A32: Average Cosine Similarity Between Questions and Answers by Party and Legislature

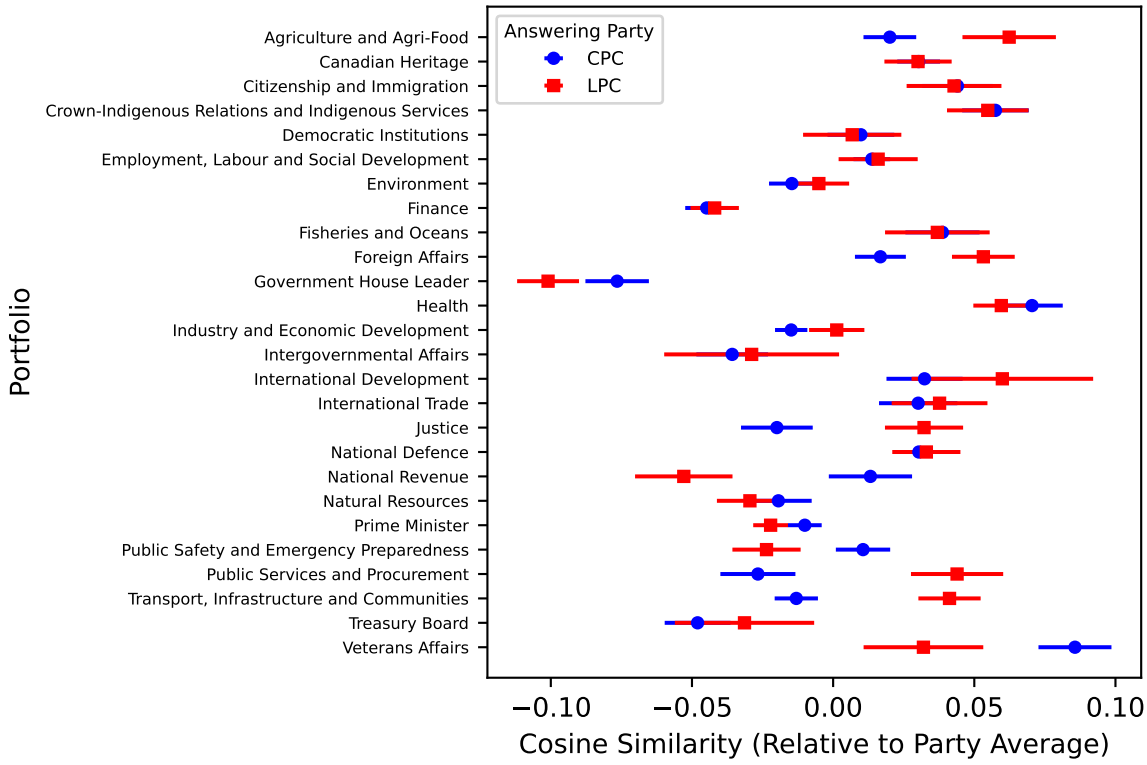


Figure A33: Average Cosine Similarity Between Questions and Answers by Party and Portfolio

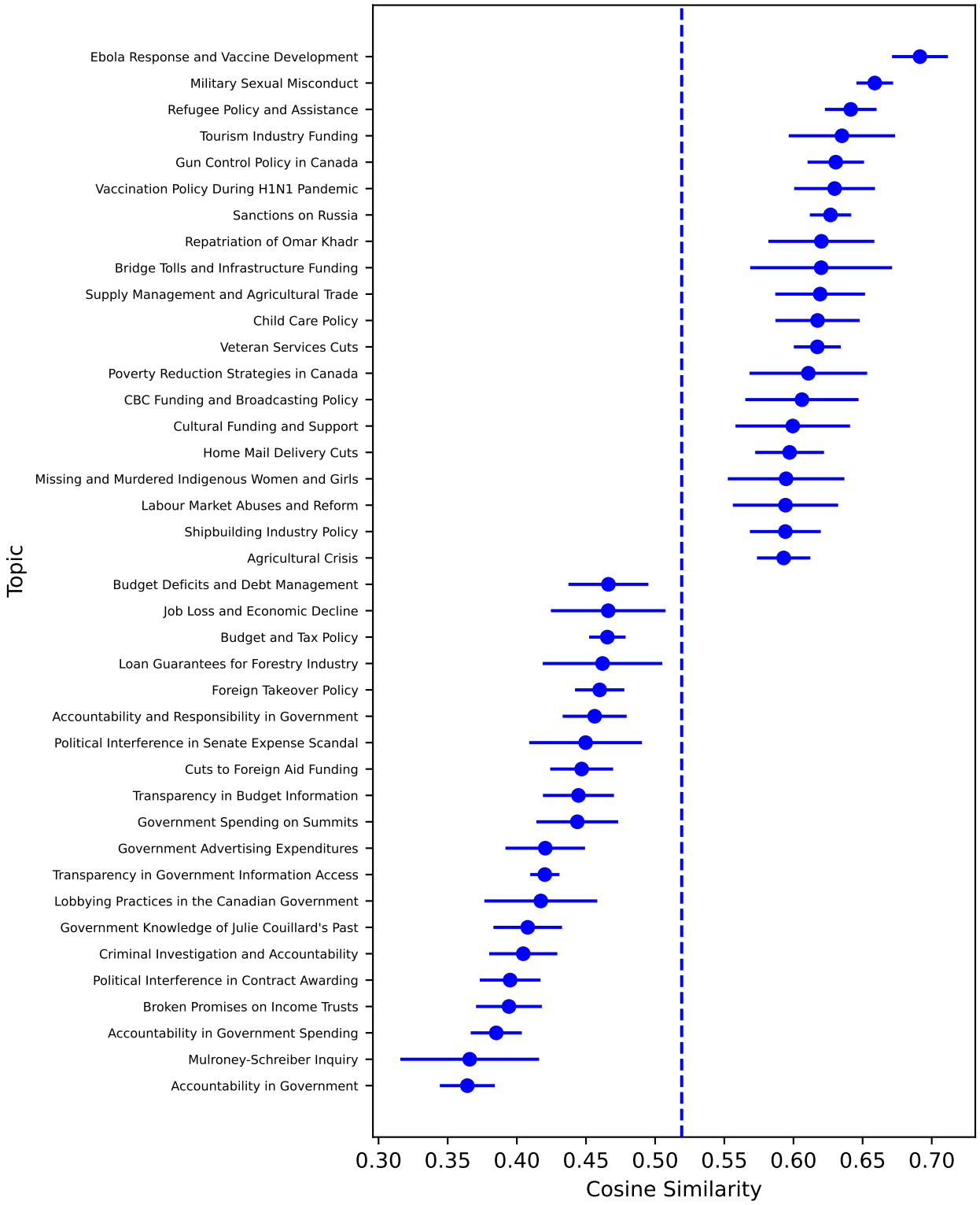


Figure A34: Topics with the 20 Lowest and Highest Average Cosine Similarity Between Questions and Answers for the Conservative Party

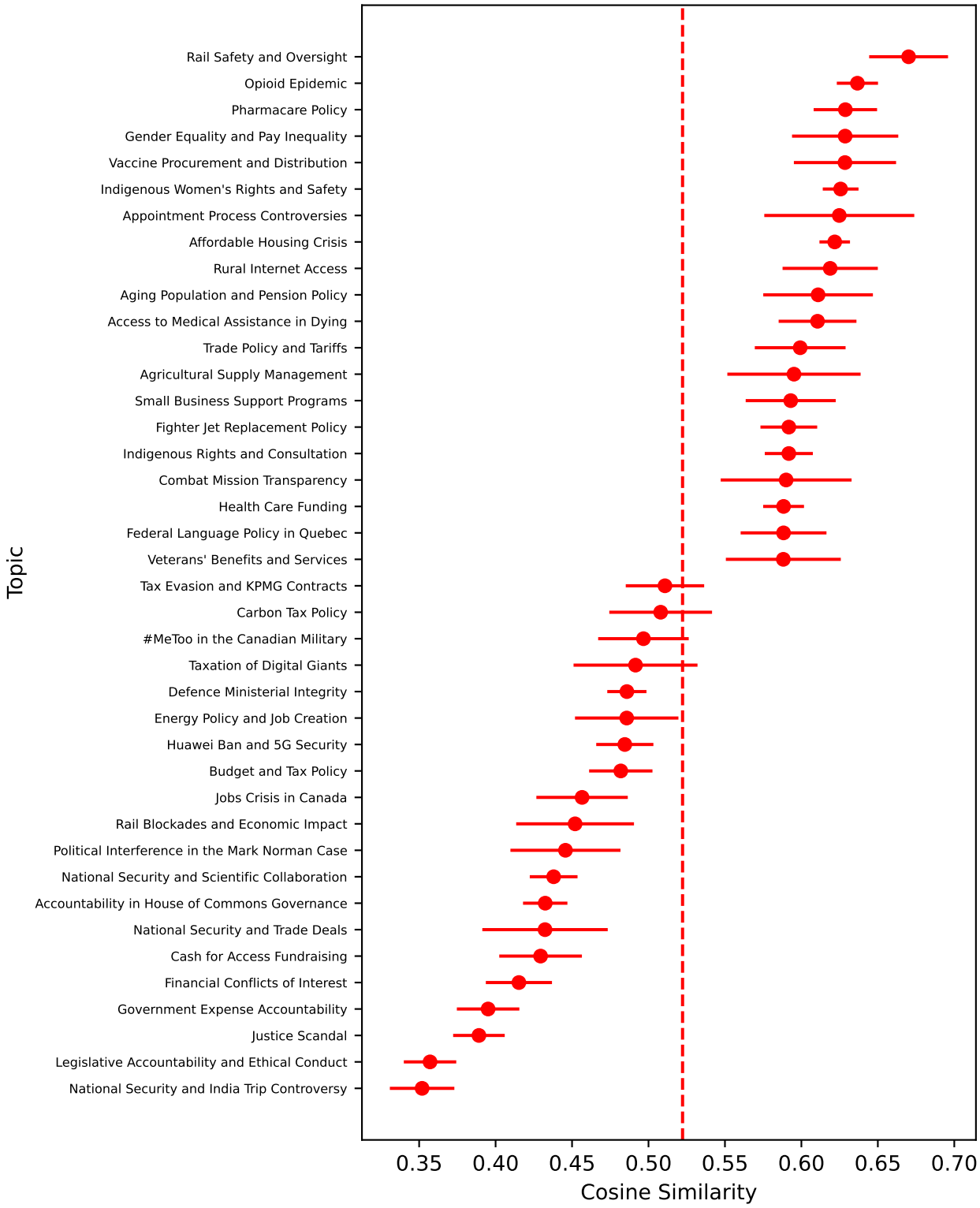


Figure A35: Topics with the 20 Lowest and Highest Average Cosine Similarity Between Questions and Answers for the Liberal Party

References

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy. 2019. "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech." *Econometrica* 87 (4): 1307–1340.

Green, Breanna, Will Hobbs, Pedro L. Rodriguez, Arthur Spirling, and Brandon M. Stewart. 2024. *Measuring Distances in High Dimensional Spaces: Why Average Group Vector Comparisons Exhibit Bias, And What to Do About it.* https://doi.org/10.31235/osf.io/g8hxt.

Loon, Austin van, Salvatore Giorgi, Robb Willer, and Johannes Eichstaedt. 2022. "Negative Associations in Word Embeddings Predict Anti-black Bias across Regions–but Only via Name Frequency." *Proceedings of the International AAAI Conference on Web and Social Media* 16 (1): 1419–1424.

Stephenson, Laura B., Allison Harell, Daniel Rubenson, and Peter John Loewen. 2020. "2019 Canadian Election Study (CES)." https://doi.org/10.7910/DVN/DUS88V.

———. 2022. "2021 Canadian Election Study (CES)." https://doi.org/10.7910/DVN/XBZHKC.