In-Context Learning for the Imputation of Survey Data

Anonymous authors
Paper under double-blind review

Abstract

We propose a Retrieval-Augmented Generation (RAG) approach for (multiple) imputation of missing values in survey data using large language models (LLMs). We evaluate various retrieval strategies, as well as performance across different missingness mechanisms (MCAR, MAR, MNAR). Using the OpinionQA dataset, we benchmark our approach against statistical imputation methods and find that using RAG with Llama-8B struggles with larger contexts and MNAR missingness. Our main contributions are (1) a thorough simulation study of LLM-based multiple imputation, and (2) a Python package with an sklearn-like API and easy local and proprietary LLM integration.

1 Introduction

Recent work has explored using large language models (LLMs) for survey data imputation, demonstrating effectiveness for some missing data patterns (Kim & Lee, 2024; Holtdirk et al., 2025; Ji et al., 2024). However, it remains unclear how well LLMs can perform multiple imputation, which involves creating several imputations to account for the uncertainty associated with missing data (Rubin, 1976) and is a well-established method in survey statistics.

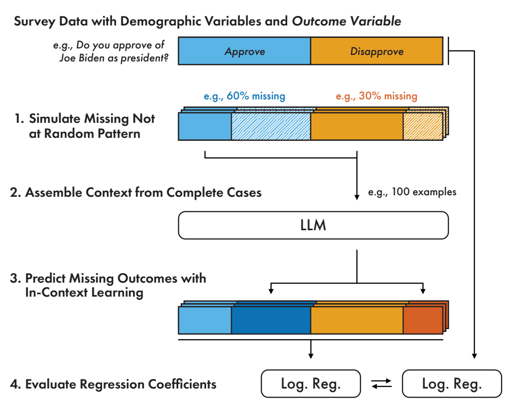

Figure 1: **Imputation pipeline using in-context learning.** We simulate different missingness patterns (1) and investigate various approaches retrieval strategies (2). We then impute the missing cases with in-context learning (3). Our evaluation is based on downstream regression coefficients (4).

We propose imputing missing survey values by prompting LLMs in a few-shot setting using retrieval augmented generation (RAG). Our approach selects auxiliary examples based on text embedding similarity to the target individual, and uses them as context for LLM-based imputation (see Figure 1). We test various selection strategies and benchmark against statistical imputation methods across different missingness mechanisms.

2 Methodology

RAG Methods We **retrieve** $n \in \{10, 100\}$ **few-shot examples** from complete cases to impute each missing case, using cosine similarity between text embeddings. We explore two **row conversion strategies**: the *list* strategy, which concatenates each variable name and value separately, and the *str* strategy, which combines all variable–value pairs into a single string. We further evaluate five **retrieval strategies**: *random_selection*, *random_selection_static_context*, *most_cosine_similar*, *least_cosine_similar*, and *cosine_diverse*. The two *random* strategies select examples at random, with *random_selection_static_context* using a fixed context across imputations. The three *cosine* strategies use embedding similarities to select either the most similar, least similar, or a diverse set of complete cases via k-medoids clustering. We also investigate **stratification** on the dependent variable for all retrieval strategies, selecting examples within each class and then shuffling the combined context. More details on the RAG methods are provided in Appendix A.1.

We conduct **two simulation studies** using the *OpinionQA* dataset from ATP Wave 92 (Santurkar et al., 2023), with demographic variables as independent variables (IVs) and Biden Approval as the dependent variable (DV). We implement three different **missingness mechanisms** (Rubin, 2018): one missing-completely-at-random (MCAR), one missing-at-random (MAR), and one missing-not-at-random (MNAR). For the **evaluation**, we fit a logistic regression on the imputed data and calculate for the beta coefficient the *bias*, *coverage*, and *width* of the confidence intervals (Van Buuren, 2018). For **study 1**, we compare the different retrieval strategies against each other. For **study 2**, we compare the RAG model against baseline methods: Complete Case, Mode Imputation, Random Sample, and MICEForest. More details on the missingness mechanisms and the study design are provided in Appendix A.2.

3 Preliminary Results

This preliminary analysis focuses on the beta for *POLIDEOLOGY*, as it is the strongest predictor for the outcome *Biden Approval*. The full results are provided in Appendix B.

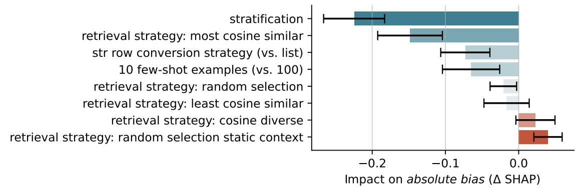

Figure 2: **Feature importance for different retrieval strategies.** Measured by SHAP values of the absolute bias for the *POLIDEOLOGY* IV on MNAR data.

Study 1 As shown in Figure 2, stratification has a substantial positive effect, likely because the Biden Approval ratio in the full data is close to uniform, resulting in a more balanced distribution of examples. The most cosine similar retrieval strategy shows the best performance, while other strategies show comparable results. This pattern suggests that the outcomes are predominantly determined by the retriever's ability to find relevant examples rather than the LLM's reasoning capabilities. Additionally, less context performs better for our examples, which might indicate that the small 7B LLM we tested struggles to effectively reason over long contexts.

Study 2 As shown in Table 3, the poor performance of LLMs on confidence interval width, being the worst in all three scenarios, shows a limitation in their handling of uncertainty. Further analysis might investigate if this can be controlled by tuning the temperature parameter of the LLM. We plan to extend this study to larger models and better-performing strategies from study 1, as well as do another study with missingness in the independent variable.

Experiment	Model	Bias	Coverage	Width
MAR	Full Data	1	1	3
	Complete Case	2	2	6
	MICE Forest	3	3	4
	Llama-3.1-8B-Instruct	**4**	**4**	**5**
	Mode	5	5	1
	Random Sample	6	5	2
MCAR	Full Data	1	2	3
	Complete Case	2	1	5
	MICE Forest	3	3	4
	Llama-3.1-8B-Instruct	**4**	**4**	**6**
	Mode	5	5	1
	Random Sample	6	5	2
MNAR	Full Data	1	1	3
	Complete Case	2	2	4
	MICE Forest	3	3	5
	Mode	4	5	1
	Llama-3.1-8B-Instruct	**5**	**4**	**6**
	Random Sample	6	5	2

Figure 3: **Model ranking for different missingness types.** Ranking of our RAG approach with 100 examples in-context, no stratification, *most_cosine_similar* retrieval strategy, and *str* row conversion strategy versus different baseline methods for the *POLIDEOLOGY* IV split by different missingness mechanisms.

References

Gordon V Cormack, Charles LA Clarke, and Stefan Buettcher. Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pp. 758–759, 2009.

Tobias Holtdirk, Dennis Assenmacher, Arnim Bleier, and Claudia Wagner. Addressing Systematic Non-response Bias with Supervised Fine-Tuning of Large Language Models: A Case Study on German Voting Behaviour, February 2025.

Junyung Ji, Jiwoo Kim, and Younghoon Kim. Predicting missing values in survey data using prompt engineering for addressing item non-response. *Future Internet*, 16(10):351, 2024.

Junsol Kim and Byungkyu Lee. AI-Augmented Surveys: Leveraging Large Language Models and Surveys for Opinion Prediction, April 2024.

Sean Lee, Aamir Shakir, Darius Koenig, and Julius Lipp. Open source strikes bread - new fluffy embeddings model, 2024. URL `https://www.mixedbread.ai/blog/mxbai-embed-large-v1`.

Nelson F Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. Lost in the middle: How language models use long contexts. *arXiv preprint arXiv:2307.03172*, 2023.

Niklas Muennighoff, Nouamane Tazi, Loïc Magne, and Nils Reimers. Mteb: Massive text embedding benchmark. *arXiv preprint arXiv:2210.07316*, 2022. doi: 10.48550/ARXIV.2210.07316. URL `https://arxiv.org/abs/2210.07316`.

Donald B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976. ISSN 00063444, 14643510. URL `http://www.jstor.org/stable/2335739`.

Donald B. Rubin. *Flexible Imputation of Missing Data*. Chapman and Hall/CRC, Boca Raton, 2nd edition, 2018. ISBN 9781138492541.

Shibani Santurkar, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. Whose Opinions Do Language Models Reflect?, March 2023.

Erich Schubert and Peter J Rousseeuw. Fast and eager k-medoids clustering: O (k) runtime improvement of the pam, clara, and clarans algorithms. *Information Systems*, 101:101804, 2021.

Stef Van Buuren. *Flexible Imputation of Missing Data, Second Edition*. Chapman and Hall/CRC, 2 edition, 2018. doi: 10.1201/9780429492259.

A Methodology Details

A.1 RAG Method Implementation

For each missing case to be imputed, **we retrieve $n \in \{10, 100\}$ few-shot examples** from the complete cases as context for the generative language model as follows. First, we compute text embeddings of all complete cases and of the missing case for which the few-shot examples are selected. We choose the `mixedbread-ai/mxbai-embed-large-v1` (Lee et al., 2024) embedding model based on the MTEB leaderboard (Muennighoff et al., 2022) for all text embeddings. We then calculate cosine similarity between the embedding of the missing case at hand and the embeddings of all complete cases.

We distinguish between two **row conversion strategies** to map rows in the dataset to strings for text embeddings: The *list* strategy concatenates the variable name with the respective value for each independent variable separately. We later combine cosine similarities for all independent variables into a single similarity measure using the Reciprocal Rank Fusion rank aggregation method (Cormack et al., 2009). We hypothesize that this strategy allows for more uniform similarity estimates across variables. In contrast, the *str* strategy combines all of these variable–value pairs into a single string to calculate the embedding.

We investigate the following **retrieval strategies** for assembling a context from complete cases: *random_selection*, *random_selection_static_context*, *most_cosine_similar*, *least_cosine_similar*, and *cosine_diverse*. Both *random* strategies select few-shot examples at random, but the *random_selection_static_context* strategy selects these examples only once for the imputation of all missing cases. This strategy has the potential to drastically reduce inference time and costs if the LLM inference provider supports automatic prefix caching.

For all three *cosine* strategies, we obtain text embedding similarities as described above. The *most_cosine_similar* strategy then selects the most similar complete cases as context for the imputation of a missing case, while the *least_cosine_similar* strategy selects the least similar cases. The *cosine_diverse* strategy uses k-medoids clustering (Schubert & Rousseeuw, 2021) to select n complete cases that are the most dissimilar to each other, while also not overemphasizing outliers.

Finally, we implement **stratification** on the dependent variable in combination with all the previously introduced retrieval strategies, that is, we independently apply any of the retrieval strategies on all rows that share a unique value of the dependent variable, before combining all retrieved cases into a context for few-shot prediction. We shuffle the combined context to avoid ordering effects (Liu et al., 2023). We hypothesize that stratification might help an LLM to more clearly distinguish between the possible values of the dependent variable, as it is presented with similar cases across all of them.

A.2 Simulation Study Design

We use the OpinionQA dataset from ATP Wave 92 (Santurkar et al., 2023), where demographic variables serve as independent variables (IVs) and Biden Approval as the dependent variable (DV). After preprocessing the data through row-wise deletion of rows with missing values, we obtain 9,511 complete cases. We conduct 10 simulations for study 1 and 100 simulations for study 2[1], where each simulation consists of a random sample of 500 participants drawn without replacement from the complete dataset. For multiple imputation models, we generate 5 imputations per simulation to account for the uncertainty associated with missing data.

Missingness Mechanisms We implement three distinct missingness mechanisms to evaluate our imputation methods. Under **Missing Completely at Random (MCAR)**, 50% of the DV values are randomly missing regardless of any observed or unobserved variables. For **Missing at Random (MAR)**, missingness depends on the IV "POLIDEOLOGY", where we give increasing missingness rates for more extreme responses: "Moderate" responses have a 20% missing rate, "Conservative" and "Liberal" responses have 40% missing rates, and "Very conservative" and "Very liberal" responses have 60% missing rates. Under **Missing Not at Random (MNAR)**, missingness depends on the DV "Biden Approval", where "Disapprove" responses have a 60% missing rate and "Approve" responses have a 30% missing rate. For study 1, we only look at MNAR, as we hypothesize that LLMs are mainly useful for imputing in MNAR scenarios. The second study uses all three missingness mechanisms.

Baseline Methods We compare our RAG-based approach against several established baseline methods for handling missing data. The **Complete Case** method performs row-wise deletion of missing values, excluding any observations with incomplete data. **Mode Imputation** replaces missing values with the most frequent value observed in the non-missing data. The **Random Sample** approach samples missing values randomly from the distribution of observed non-missing values. Finally, we include **MICEForest**, which implements Multiple Imputation by Chained Equations using Random Forest models to generate multiple plausible imputations for each missing value.

[1] 500-1000 simulations are typical in imputation methods research, however, given the number of parameters we are exploring and the computational cost of LLMs we have reduced this for now.

B Results Full Tables

		coef		bias		coverage		width	
		mean	std	mean	std	mean	std	mean	std
variable	model								
AGE	Full Data	-0.095	0.127	0.004	0.127	0.930	0.256	0.473	0.022
	Complete Case	-0.114	0.190	0.023	0.190	0.940	0.239	0.684	0.048
	Mode Imputation	-0.055	0.131	-0.036	0.131	0.940	0.239	0.482	0.018
	Random Sample	-0.043	0.077	-0.048	0.077	0.990	0.100	0.540	0.094
	MICE Forest	-0.111	0.219	0.020	0.219	0.840	0.368	0.585	0.071
	Llama-3.1-8B-Instruct	-0.079	0.163	-0.013	0.163	0.940	0.239	0.620	0.075
CITIZEN	Full Data	-0.300	0.163	0.032	0.163	0.918	0.277	0.559	0.114
	Complete Case	-0.249	0.218	-0.019	0.218	0.928	0.261	0.766	0.159
	Mode Imputation	-0.165	0.139	-0.103	0.139	0.831	0.377	0.620	0.146
	Random Sample	-0.106	0.082	-0.162	0.082	0.818	0.388	0.569	0.154
	MICE Forest	-0.239	0.234	-0.029	0.234	0.901	0.300	0.699	0.180
	Llama-3.1-8B-Instruct	-0.054	0.195	-0.214	0.195	0.707	0.457	0.708	0.208
EDUCATION	Full Data	-0.097	0.131	0.018	0.131	0.950	0.219	0.509	0.022
	Complete Case	-0.108	0.176	0.028	0.176	0.970	0.171	0.736	0.053
	Mode Imputation	-0.058	0.123	-0.021	0.123	0.950	0.219	0.515	0.018
	Random Sample	-0.037	0.075	-0.043	0.075	1.000	0.000	0.571	0.108
	MICE Forest	-0.122	0.199	0.042	0.199	0.890	0.314	0.626	0.073
	Llama-3.1-8B-Instruct	-0.081	0.131	0.001	0.131	0.980	0.141	0.676	0.092
INCOME	Full Data	0.074	0.121	-0.006	0.121	0.970	0.171	0.507	0.025
	Complete Case	0.077	0.169	-0.009	0.169	0.980	0.141	0.732	0.055
	Mode Imputation	0.037	0.131	0.031	0.131	0.940	0.239	0.511	0.020
	Random Sample	0.030	0.079	0.038	0.079	1.000	0.000	0.580	0.105
	MICE Forest	0.079	0.215	-0.011	0.215	0.880	0.327	0.636	0.087
	Llama-3.1-8B-Instruct	0.093	0.171	-0.025	0.171	0.970	0.171	0.681	0.103
POLIDEOLOGY	Full Data	-1.833	0.235	0.026	0.235	0.820	0.386	0.703	0.069
	Complete Case	-1.882	0.351	0.075	0.351	0.930	0.256	1.022	0.162
	Mode Imputation	-0.988	0.150	-0.818	0.150	0.000	0.000	0.584	0.040
	Random Sample	-0.644	0.086	-1.162	0.086	0.000	0.000	0.589	0.109
	MICE Forest	-2.228	0.420	0.421	0.420	0.650	0.479	1.015	0.244
	Llama-3.1-8B-Instruct	-2.563	0.450	0.757	0.450	0.310	0.465	1.056	0.177
RELIGATTEND	Full Data	-0.028	0.121	-0.010	0.121	0.940	0.239	0.484	0.021
	Complete Case	-0.028	0.169	-0.011	0.169	0.960	0.197	0.695	0.051
	Mode Imputation	-0.021	0.123	-0.018	0.123	0.950	0.219	0.487	0.034
	Random Sample	-0.007	0.079	-0.032	0.079	1.000	0.000	0.533	0.083
	MICE Forest	-0.048	0.198	0.009	0.198	0.870	0.338	0.594	0.069
	Llama-3.1-8B-Instruct	-0.041	0.149	0.003	0.149	0.980	0.141	0.627	0.069

Figure 4: **Study 2 – MCAR.** Coefficient, bias, and confidence interval coverage and width for our RAG approach and various baseline methods for the all IVs on MCAR data. This retriever is configured with 100 examples in-context, no stratification, most cosine similar retrieval strategy, and str row conversion strategy.

variable	model	coef mean	coef std	bias mean	bias std	coverage mean	coverage std	width mean	width std
AGE	Full Data	-0.104	0.130	0.012	0.130	0.930	0.256	0.475	0.019
	Complete Case	-0.105	0.166	0.013	0.166	0.930	0.256	0.583	0.031
	Mode Imputation	-0.042	0.121	-0.049	0.121	0.890	0.314	0.442	0.013
	Random Sample	-0.032	0.089	-0.059	0.089	0.960	0.197	0.491	0.069
	MICE Forest	-0.102	0.177	0.011	0.177	0.890	0.314	0.545	0.053
	Llama-3.1-8B-Instruct	-0.093	0.160	0.002	0.160	0.930	0.256	0.566	0.041
CITIZEN	Full Data	-0.322	0.182	0.054	0.182	0.910	0.288	0.587	0.147
	Complete Case	-0.280	0.179	0.012	0.179	0.947	0.226	0.683	0.162
	Mode Imputation	-0.195	0.153	-0.073	0.153	0.840	0.368	0.601	0.154
	Random Sample	-0.149	0.091	-0.119	0.091	0.878	0.329	0.551	0.150
	MICE Forest	-0.271	0.188	0.003	0.188	0.938	0.243	0.656	0.178
	Llama-3.1-8B-Instruct	-0.170	0.173	-0.098	0.173	0.848	0.360	0.637	0.164
EDUCATION	Full Data	-0.081	0.135	0.001	0.135	0.960	0.197	0.513	0.022
	Complete Case	-0.087	0.164	0.007	0.164	0.970	0.171	0.631	0.034
	Mode Imputation	-0.066	0.135	-0.014	0.135	0.940	0.239	0.477	0.016
	Random Sample	-0.036	0.086	-0.044	0.086	1.000	0.000	0.526	0.074
	MICE Forest	-0.090	0.171	0.010	0.171	0.930	0.256	0.589	0.050
	Llama-3.1-8B-Instruct	-0.083	0.143	0.004	0.143	0.980	0.141	0.618	0.058
INCOME	Full Data	0.058	0.127	0.010	0.127	0.940	0.239	0.513	0.021
	Complete Case	0.054	0.160	0.014	0.160	0.970	0.171	0.631	0.036
	Mode Imputation	0.028	0.129	0.040	0.129	0.900	0.302	0.476	0.016
	Random Sample	0.037	0.091	0.031	0.091	0.990	0.100	0.522	0.064
	MICE Forest	0.053	0.173	0.015	0.173	0.930	0.256	0.594	0.061
	Llama-3.1-8B-Instruct	0.075	0.149	-0.007	0.149	0.980	0.141	0.616	0.057
POLIDEOLOGY	Full Data	-1.906	0.213	0.099	0.213	0.920	0.273	0.724	0.064
	Complete Case	-2.073	0.324	0.267	0.324	0.850	0.359	1.028	0.124
	Mode Imputation	-0.879	0.112	-0.927	0.112	0.000	0.000	0.527	0.029
	Random Sample	-0.686	0.087	-1.121	0.087	0.000	0.000	0.622	0.120
	MICE Forest	-2.188	0.374	0.381	0.374	0.690	0.465	1.004	0.180
	Llama-3.1-8B-Instruct	-2.587	0.412	0.780	0.412	0.180	0.386	1.023	0.133
RELIGATTEND	Full Data	-0.049	0.135	0.010	0.135	0.920	0.273	0.489	0.020
	Complete Case	-0.055	0.170	0.017	0.170	0.940	0.239	0.601	0.032
	Mode Imputation	-0.053	0.137	0.014	0.137	0.920	0.273	0.454	0.018
	Random Sample	-0.020	0.095	-0.019	0.095	0.990	0.100	0.513	0.075
	MICE Forest	-0.054	0.186	0.015	0.186	0.890	0.314	0.555	0.048
	Llama-3.1-8B-Instruct	-0.059	0.152	0.020	0.152	0.940	0.239	0.580	0.048

Figure 5: **Study 2 – MAR.** Coefficient, bias, and confidence interval coverage and width for our RAG approach and various baseline methods for the all IVs on MAR data. This retriever is configured with 100 examples in-context, no stratification, most cosine similar retrieval strategy, and str row conversion strategy.

		coef		bias		coverage		width	
		mean	std	mean	std	mean	std	mean	std
variable	model								
AGE	Full Data	-0.112	0.120	0.021	0.120	0.940	0.239	0.473	0.025
	Complete Case	-0.063	0.176	-0.029	0.176	0.960	0.197	0.680	0.054
	Mode Imputation	-0.043	0.137	-0.048	0.137	0.940	0.239	0.512	0.022
	Random Sample	-0.046	0.074	-0.046	0.074	1.000	0.000	0.529	0.089
	MICE Forest	-0.015	0.181	-0.076	0.181	0.850	0.359	0.600	0.082
	Llama-3.1-8B-Instruct	-0.017	0.159	-0.075	0.159	0.940	0.239	0.623	0.062
CITIZEN	Full Data	-0.327	0.177	0.059	0.177	0.937	0.245	0.583	0.139
	Complete Case	-0.228	0.151	-0.040	0.151	0.971	0.169	0.737	0.138
	Mode Imputation	-0.113	0.116	-0.155	0.116	0.855	0.355	0.647	0.133
	Random Sample	-0.141	0.096	-0.127	0.096	0.916	0.279	0.623	0.195
	MICE Forest	-0.253	0.180	-0.015	0.180	0.938	0.244	0.750	0.178
	Llama-3.1-8B-Instruct	-0.147	0.192	-0.120	0.192	0.848	0.361	0.750	0.230
EDUCATION	Full Data	-0.103	0.123	0.024	0.123	0.960	0.197	0.509	0.027
	Complete Case	-0.085	0.187	0.005	0.187	0.950	0.219	0.728	0.053
	Mode Imputation	-0.063	0.153	-0.017	0.153	0.930	0.256	0.550	0.023
	Random Sample	-0.038	0.081	-0.042	0.081	1.000	0.000	0.582	0.111
	MICE Forest	-0.069	0.203	-0.011	0.203	0.880	0.327	0.636	0.075
	Llama-3.1-8B-Instruct	-0.072	0.139	-0.008	0.139	0.970	0.171	0.666	0.071
INCOME	Full Data	0.088	0.121	-0.020	0.121	0.970	0.171	0.507	0.029
	Complete Case	0.119	0.188	-0.051	0.188	0.960	0.197	0.728	0.061
	Mode Imputation	0.054	0.136	0.014	0.136	0.950	0.219	0.545	0.025
	Random Sample	0.032	0.081	0.036	0.081	1.000	0.000	0.582	0.109
	MICE Forest	0.147	0.217	-0.079	0.217	0.880	0.327	0.649	0.091
	Llama-3.1-8B-Instruct	0.134	0.164	-0.066	0.164	0.950	0.219	0.684	0.090
POLIDEOLOGY	Full Data	-1.849	0.265	0.042	0.265	0.820	0.386	0.707	0.077
	Complete Case	-1.930	0.409	0.123	0.409	0.800	0.402	1.026	0.179
	Mode Imputation	-0.932	0.170	-0.875	0.170	0.000	0.000	0.608	0.072
	Random Sample	-0.755	0.101	-1.052	0.101	0.000	0.000	0.643	0.130
	MICE Forest	-2.275	0.495	0.468	0.495	0.580	0.496	1.053	0.260
	Llama-3.1-8B-Instruct	-2.713	0.553	0.907	0.553	0.130	0.338	1.067	0.193
RELIGATTEND	Full Data	-0.041	0.120	0.002	0.120	0.950	0.219	0.484	0.022
	Complete Case	-0.039	0.167	0.000	0.167	0.980	0.141	0.692	0.049
	Mode Imputation	-0.026	0.140	-0.013	0.140	0.930	0.256	0.519	0.022
	Random Sample	-0.016	0.078	-0.023	0.078	1.000	0.000	0.572	0.113
	MICE Forest	-0.041	0.193	0.002	0.193	0.860	0.349	0.592	0.066
	Llama-3.1-8B-Instruct	-0.035	0.147	-0.004	0.147	0.990	0.100	0.625	0.067

Figure 6: **Study 2 – MNAR.** Coefficient, bias, and confidence interval coverage and width for our RAG approach and various baseline methods for the all IVs on MNAR data. This retriever is configured with 100 examples in-context, no stratification, most cosine similar retrieval strategy, and str row conversion strategy.

few-shot examples	retrieval strategy	row conversion strategy	stratification	coef mean	coef std	bias mean	bias std	coverage mean	coverage std	width mean	width std
100	random_selection_static_context	row_to_text_str	True	-0.023	0.121	-0.068	0.121	0.800	0.422	0.443	0.031
			False	0.006	0.159	-0.097	0.159	1.000	0.000	0.638	0.060
		row_to_text_list	True	-0.030	0.108	-0.062	0.108	0.900	0.316	0.449	0.035
			False	0.000	0.161	-0.092	0.161	0.900	0.316	0.626	0.047
	random_selection	row_to_text_str	True	-0.032	0.099	-0.060	0.099	1.000	0.000	0.441	0.024
			False	0.023	0.144	-0.115	0.144	1.000	0.000	0.642	0.064
		row_to_text_list	True	-0.032	0.103	-0.059	0.103	0.900	0.316	0.443	0.030
			False	-0.011	0.142	-0.080	0.142	1.000	0.000	0.620	0.069
	most_cosine_similar	row_to_text_str	True	0.057	0.104	-0.148	0.104	0.900	0.316	0.523	0.040
			False	0.014	0.150	-0.105	0.150	1.000	0.000	0.651	0.115
		row_to_text_list	True	0.006	0.083	-0.098	0.083	0.900	0.316	0.488	0.028
			False	0.029	0.154	-0.121	0.154	1.000	0.000	0.632	0.053
	least_cosine_similar	row_to_text_str	True	0.041	0.061	-0.132	0.061	0.900	0.316	0.497	0.040
			False	-0.001	0.107	-0.090	0.107	0.900	0.316	0.620	0.049
		row_to_text_list	True	-0.010	0.063	-0.081	0.063	1.000	0.000	0.440	0.030
			False	0.027	0.100	-0.119	0.100	1.000	0.000	0.642	0.061
	cosine_diverse	row_to_text_str	True	-0.032	0.105	-0.059	0.105	0.900	0.316	0.436	0.034
			False	-0.008	0.111	-0.083	0.111	1.000	0.000	0.605	0.062
		row_to_text_list	True	-0.028	0.110	-0.064	0.110	0.900	0.316	0.448	0.038
			False	0.000	0.141	-0.091	0.141	1.000	0.000	0.673	0.092
10	random_selection_static_context	row_to_text_str	True	0.016	0.149	-0.107	0.149	0.900	0.316	0.573	0.056
			False	0.011	0.114	-0.102	0.114	1.000	0.000	0.590	0.047
		row_to_text_list	True	0.018	0.145	-0.110	0.145	0.900	0.316	0.596	0.067
			False	0.015	0.117	-0.107	0.117	1.000	0.000	0.584	0.063
	random_selection	row_to_text_str	True	-0.019	0.130	-0.072	0.130	1.000	0.000	0.574	0.060
			False	0.010	0.103	-0.101	0.103	0.900	0.316	0.564	0.041
		row_to_text_list	True	-0.027	0.097	-0.065	0.097	1.000	0.000	0.570	0.065
			False	0.042	0.109	-0.134	0.109	1.000	0.000	0.600	0.075
	most_cosine_similar	row_to_text_str	True	-0.019	0.144	-0.072	0.144	1.000	0.000	0.590	0.079
			False	0.041	0.134	-0.132	0.134	0.900	0.316	0.579	0.074
		row_to_text_list	True	-0.018	0.158	-0.073	0.158	1.000	0.000	0.581	0.045
			False	0.007	0.140	-0.098	0.140	0.900	0.316	0.560	0.045
	least_cosine_similar	row_to_text_str	True	-0.003	0.092	-0.089	0.092	1.000	0.000	0.601	0.050
			False	0.017	0.120	-0.108	0.120	1.000	0.000	0.576	0.057
		row_to_text_list	True	0.041	0.104	-0.132	0.104	1.000	0.000	0.618	0.099
			False	0.066	0.102	-0.157	0.102	0.900	0.316	0.593	0.050
	cosine_diverse	row_to_text_str	True	-0.006	0.094	-0.085	0.094	1.000	0.000	0.580	0.061
			False	0.029	0.128	-0.121	0.128	1.000	0.000	0.598	0.062
		row_to_text_list	True	0.028	0.132	-0.120	0.132	0.900	0.316	0.553	0.059
			False	-0.003	0.113	-0.088	0.113	1.000	0.000	0.582	0.052

Table 1: **Study 1 – AGE.** Simulation results for "Llama-3.1-8B-Instruct" with various retriever configurations for the IV AGE.

few-shot examples	retrieval strategy	row conversion strategy	stratification	coef mean	coef std	bias mean	bias std	coverage mean	coverage std	width mean	width std
100	random_selection_static_context	row_to_text_str	True	-0.197	0.095	-0.071	0.095	1.000	0.000	0.453	0.083
			False	-0.141	0.144	-0.127	0.144	0.889	0.333	0.768	0.111
		row_to_text_list	True	-0.194	0.096	-0.074	0.096	1.000	0.000	0.460	0.083
			False	-0.141	0.150	-0.127	0.150	0.889	0.333	0.730	0.165
	random_selection	row_to_text_str	True	-0.191	0.087	-0.077	0.087	1.000	0.000	0.454	0.083
			False	-0.198	0.225	-0.070	0.225	0.889	0.333	0.822	0.164
		row_to_text_list	True	-0.181	0.096	-0.087	0.096	1.000	0.000	0.440	0.087
			False	-0.154	0.123	-0.114	0.123	0.889	0.333	0.818	0.199
	most_cosine_similar	row_to_text_str	True	-0.112	0.091	-0.156	0.091	0.900	0.316	0.543	0.094
			False	-0.177	0.171	-0.091	0.171	0.889	0.333	0.908	0.293
		row_to_text_list	True	-0.175	0.081	-0.093	0.081	1.000	0.000	0.492	0.085
			False	-0.243	0.186	-0.025	0.186	0.889	0.333	0.786	0.151
	least_cosine_similar	row_to_text_str	True	-0.203	0.120	-0.065	0.120	1.000	0.000	0.512	0.106
			False	-0.219	0.200	-0.049	0.200	1.000	0.000	0.731	0.152
		row_to_text_list	True	-0.168	0.154	-0.100	0.154	0.900	0.316	0.450	0.076
			False	-0.112	0.169	-0.156	0.169	0.778	0.441	0.710	0.151
	cosine_diverse	row_to_text_str	True	-0.185	0.097	-0.083	0.097	1.000	0.000	0.438	0.080
			False	-0.172	0.161	-0.096	0.161	0.889	0.333	0.841	0.264
		row_to_text_list	True	-0.197	0.091	-0.071	0.091	1.000	0.000	0.456	0.080
			False	-0.151	0.125	-0.116	0.125	0.889	0.333	0.780	0.156
10	random_selection_static_context	row_to_text_str	True	0.002	0.148	-0.270	0.148	0.500	0.527	0.534	0.084
			False	0.038	0.132	-0.306	0.132	0.200	0.422	0.549	0.119
		row_to_text_list	True	0.003	0.153	-0.271	0.153	0.500	0.527	0.520	0.055
			False	0.037	0.119	-0.305	0.119	0.300	0.483	0.553	0.136
	random_selection	row_to_text_str	True	-0.021	0.130	-0.247	0.130	0.400	0.516	0.527	0.073
			False	-0.003	0.108	-0.265	0.108	0.600	0.516	0.564	0.133
		row_to_text_list	True	-0.028	0.095	-0.240	0.095	0.500	0.527	0.526	0.067
			False	-0.050	0.119	-0.218	0.119	0.600	0.516	0.609	0.155
	most_cosine_similar	row_to_text_str	True	-0.006	0.120	-0.262	0.120	0.600	0.516	0.518	0.043
			False	-0.032	0.140	-0.236	0.140	0.700	0.483	0.584	0.111
		row_to_text_list	True	-0.039	0.138	-0.229	0.138	0.600	0.516	0.625	0.212
			False	-0.105	0.172	-0.163	0.172	0.700	0.483	0.768	0.259
	least_cosine_similar	row_to_text_str	True	-0.077	0.107	-0.191	0.107	0.900	0.316	0.636	0.240
			False	-0.055	0.178	-0.213	0.178	0.500	0.527	0.627	0.240
		row_to_text_list	True	-0.021	0.078	-0.247	0.078	0.500	0.527	0.535	0.077
			False	-0.047	0.141	-0.221	0.141	0.700	0.483	0.615	0.117
	cosine_diverse	row_to_text_str	True	-0.017	0.094	-0.251	0.094	0.400	0.516	0.495	0.064
			False	0.063	0.115	-0.331	0.115	0.200	0.422	0.550	0.087
		row_to_text_list	True	0.009	0.151	-0.277	0.151	0.500	0.527	0.506	0.026
			False	0.041	0.116	-0.309	0.116	0.300	0.483	0.579	0.124

Table 2: **Study 1 – CITIZEN.** Simulation results for "Llama-3.1-8B-Instruct" with various retriever configurations for the IV CITIZEN.

few-shot examples	retrieval strategy	row conversion strategy	stratification	coef mean	coef std	bias mean	bias std	coverage mean	coverage std	width mean	width std
100	random_selection_static_context	row_to_text_str	True	0.023	0.149	-0.102	0.149	0.800	0.422	0.489	0.042
			False	-0.030	0.116	-0.049	0.116	1.000	0.000	0.690	0.069
		row_to_text_list	True	0.017	0.146	-0.096	0.146	0.800	0.422	0.491	0.047
			False	-0.034	0.129	-0.046	0.129	1.000	0.000	0.655	0.052
	random_selection	row_to_text_str	True	0.015	0.158	-0.095	0.158	0.800	0.422	0.484	0.029
			False	-0.013	0.085	-0.067	0.085	1.000	0.000	0.666	0.052
		row_to_text_list	True	0.024	0.145	-0.104	0.145	0.700	0.483	0.486	0.034
			False	-0.052	0.134	-0.028	0.134	1.000	0.000	0.684	0.076
	most_cosine_similar	row_to_text_str	True	-0.023	0.152	-0.057	0.152	1.000	0.000	0.573	0.049
			False	-0.057	0.133	-0.023	0.133	1.000	0.000	0.675	0.049
		row_to_text_list	True	0.022	0.157	-0.102	0.157	0.800	0.422	0.522	0.024
			False	-0.047	0.132	-0.033	0.132	1.000	0.000	0.672	0.064
	least_cosine_similar	row_to_text_str	True	0.044	0.155	-0.123	0.155	0.800	0.422	0.541	0.060
			False	-0.049	0.095	-0.031	0.095	1.000	0.000	0.666	0.055
		row_to_text_list	True	0.029	0.138	-0.109	0.138	0.700	0.483	0.480	0.030
			False	-0.032	0.110	-0.047	0.110	1.000	0.000	0.674	0.058
	cosine_diverse	row_to_text_str	True	0.023	0.164	-0.102	0.164	0.700	0.483	0.477	0.030
			False	-0.005	0.132	-0.074	0.132	1.000	0.000	0.676	0.037
		row_to_text_list	True	0.029	0.146	-0.109	0.146	0.800	0.422	0.494	0.049
			False	-0.035	0.101	-0.045	0.101	1.000	0.000	0.684	0.068
10	random_selection_static_context	row_to_text_str	True	-0.054	0.129	-0.026	0.129	0.900	0.316	0.600	0.058
			False	-0.085	0.141	0.006	0.141	1.000	0.000	0.623	0.063
		row_to_text_list	True	-0.046	0.131	-0.033	0.131	0.900	0.316	0.596	0.047
			False	-0.087	0.133	0.008	0.133	1.000	0.000	0.616	0.054
	random_selection	row_to_text_str	True	-0.078	0.131	-0.002	0.131	1.000	0.000	0.601	0.054
			False	-0.050	0.133	-0.030	0.133	1.000	0.000	0.624	0.057
		row_to_text_list	True	-0.052	0.122	-0.028	0.122	1.000	0.000	0.606	0.046
			False	-0.057	0.113	-0.023	0.113	1.000	0.000	0.613	0.048
	most_cosine_similar	row_to_text_str	True	0.011	0.154	-0.091	0.154	0.900	0.316	0.641	0.097
			False	0.021	0.142	-0.101	0.142	0.900	0.316	0.632	0.036
		row_to_text_list	True	-0.010	0.128	-0.069	0.128	1.000	0.000	0.615	0.065
			False	-0.014	0.153	-0.066	0.153	1.000	0.000	0.638	0.042
	least_cosine_similar	row_to_text_str	True	-0.030	0.147	-0.049	0.147	1.000	0.000	0.636	0.053
			False	-0.051	0.134	-0.029	0.134	1.000	0.000	0.619	0.030
		row_to_text_list	True	-0.072	0.146	-0.008	0.146	1.000	0.000	0.594	0.041
			False	-0.102	0.117	0.022	0.117	1.000	0.000	0.621	0.033
	cosine_diverse	row_to_text_str	True	-0.049	0.134	-0.031	0.134	1.000	0.000	0.590	0.050
			False	-0.094	0.139	0.015	0.139	1.000	0.000	0.632	0.050
		row_to_text_list	True	-0.043	0.139	-0.037	0.139	0.900	0.316	0.609	0.067
			False	-0.070	0.148	-0.010	0.148	1.000	0.000	0.602	0.032

Table 3: **Study 1 – EDUCATION.** Simulation results for ”Llama-3.1-8B-Instruct” with various retriever configurations for the IV EDUCATION.

few-shot examples	retrieval strategy	row conversion strategy	stratification	coef mean	coef std	bias mean	bias std	coverage mean	coverage std	width mean	width std
100	random_selection_static_context	row_to_text_str	True	-0.020	0.137	0.088	0.137	0.700	0.483	0.485	0.048
			False	0.147	0.186	-0.079	0.186	0.900	0.316	0.703	0.040
		row_to_text_list	True	-0.021	0.128	0.089	0.128	0.800	0.422	0.485	0.048
			False	0.146	0.198	-0.078	0.198	0.900	0.316	0.682	0.058
	random_selection	row_to_text_str	True	-0.011	0.129	0.079	0.129	0.800	0.422	0.481	0.031
			False	0.119	0.183	-0.051	0.183	0.900	0.316	0.672	0.043
		row_to_text_list	True	-0.018	0.103	0.086	0.103	0.900	0.316	0.482	0.034
			False	0.125	0.200	-0.057	0.200	0.900	0.316	0.708	0.069
	most_cosine_similar	row_to_text_str	True	0.061	0.139	0.007	0.139	1.000	0.000	0.574	0.046
			False	0.148	0.181	-0.080	0.181	0.800	0.422	0.715	0.087
		row_to_text_list	True	0.033	0.128	0.035	0.128	1.000	0.000	0.518	0.028
			False	0.122	0.189	-0.054	0.189	1.000	0.000	0.719	0.094
	least_cosine_similar	row_to_text_str	True	0.076	0.118	-0.008	0.118	1.000	0.000	0.547	0.065
			False	0.089	0.130	-0.021	0.130	1.000	0.000	0.659	0.044
		row_to_text_list	True	0.021	0.113	0.047	0.113	0.900	0.316	0.483	0.032
			False	0.084	0.124	-0.016	0.124	1.000	0.000	0.669	0.061
	cosine_diverse	row_to_text_str	True	-0.009	0.122	0.077	0.122	0.800	0.422	0.470	0.037
			False	0.155	0.178	-0.087	0.178	0.900	0.316	0.683	0.080
		row_to_text_list	True	-0.020	0.125	0.088	0.125	0.800	0.422	0.485	0.048
			False	0.138	0.198	-0.070	0.198	0.900	0.316	0.676	0.069
10	random_selection_static_context	row_to_text_str	True	-0.126	0.208	0.194	0.208	0.600	0.516	0.626	0.044
			False	-0.039	0.215	0.107	0.215	0.900	0.316	0.635	0.061
		row_to_text_list	True	-0.130	0.225	0.198	0.225	0.700	0.483	0.643	0.080
			False	-0.045	0.211	0.113	0.211	0.900	0.316	0.634	0.048
	random_selection	row_to_text_str	True	-0.058	0.162	0.126	0.162	0.900	0.316	0.608	0.047
			False	-0.028	0.151	0.096	0.151	0.900	0.316	0.622	0.047
		row_to_text_list	True	-0.056	0.146	0.124	0.146	0.900	0.316	0.617	0.029
			False	-0.044	0.140	0.112	0.140	1.000	0.000	0.620	0.044
	most_cosine_similar	row_to_text_str	True	-0.078	0.132	0.146	0.132	1.000	0.000	0.612	0.075
			False	-0.045	0.191	0.113	0.191	0.900	0.316	0.632	0.039
		row_to_text_list	True	-0.077	0.170	0.145	0.170	0.800	0.422	0.621	0.074
			False	-0.019	0.166	0.087	0.166	1.000	0.000	0.632	0.047
	least_cosine_similar	row_to_text_str	True	-0.039	0.103	0.107	0.103	1.000	0.000	0.639	0.081
			False	-0.000	0.109	0.068	0.109	1.000	0.000	0.646	0.074
		row_to_text_list	True	-0.050	0.113	0.118	0.113	0.900	0.316	0.604	0.042
			False	0.023	0.142	0.045	0.142	1.000	0.000	0.645	0.053
	cosine_diverse	row_to_text_str	True	-0.130	0.153	0.198	0.153	0.600	0.516	0.588	0.060
			False	0.010	0.169	0.058	0.169	1.000	0.000	0.638	0.044
		row_to_text_list	True	-0.121	0.209	0.189	0.209	0.600	0.516	0.637	0.063
			False	-0.053	0.217	0.121	0.217	0.800	0.422	0.664	0.109

Table 4: **Study 1 – INCOME.** Simulation results for ”Llama-3.1-8B-Instruct” with various retriever configurations for the IV INCOME.

few-shot examples	retrieval strategy	row conversion strategy	stratification	coef mean	coef std	bias mean	bias std	coverage mean	coverage std	width mean	width std
100	random_selection_static_context	row_to_text_str	True	-1.197	0.407	-0.610	0.407	0.100	0.316	0.595	0.127
			False	-2.657	0.412	0.850	0.412	0.100	0.316	1.045	0.147
		row_to_text_list	True	-1.200	0.420	-0.607	0.420	0.100	0.316	0.633	0.175
			False	-2.653	0.401	0.847	0.401	0.100	0.316	1.042	0.145
	random_selection	row_to_text_str	True	-1.136	0.185	-0.670	0.185	0.000	0.000	0.575	0.053
			False	-2.671	0.307	0.865	0.307	0.100	0.316	1.114	0.138
		row_to_text_list	True	-1.133	0.189	-0.674	0.189	0.000	0.000	0.573	0.059
			False	-2.705	0.320	0.898	0.320	0.100	0.316	1.089	0.113
	most_cosine_similar	row_to_text_str	True	-1.826	0.398	0.020	0.398	0.800	0.422	0.863	0.162
			False	-2.536	0.324	0.729	0.324	0.200	0.422	1.197	0.197
		row_to_text_list	True	-1.435	0.306	-0.372	0.306	0.500	0.527	0.735	0.188
			False	-2.728	0.299	0.922	0.299	0.100	0.316	1.076	0.129
	least_cosine_similar	row_to_text_str	True	-1.495	0.207	-0.312	0.207	0.700	0.483	0.688	0.096
			False	-2.785	0.358	0.979	0.358	0.100	0.316	1.111	0.118
		row_to_text_list	True	-1.128	0.191	-0.678	0.191	0.000	0.000	0.568	0.060
			False	-2.684	0.382	0.878	0.382	0.200	0.422	1.080	0.142
	cosine_diverse	row_to_text_str	True	-1.078	0.207	-0.729	0.207	0.100	0.316	0.545	0.069
			False	-2.514	0.346	0.707	0.346	0.300	0.483	1.053	0.083
		row_to_text_list	True	-1.204	0.416	-0.602	0.416	0.100	0.316	0.587	0.120
			False	-2.676	0.414	0.869	0.414	0.200	0.422	1.117	0.143
10	random_selection_static_context	row_to_text_str	True	-2.412	0.481	0.606	0.481	0.300	0.483	0.994	0.186
			False	-2.600	0.342	0.794	0.342	0.200	0.422	1.044	0.145
		row_to_text_list	True	-2.423	0.513	0.617	0.513	0.200	0.422	0.991	0.177
			False	-2.635	0.349	0.828	0.349	0.100	0.316	1.006	0.125
	random_selection	row_to_text_str	True	-2.325	0.262	0.519	0.262	0.400	0.516	1.039	0.183
			False	-2.507	0.371	0.701	0.371	0.300	0.483	1.023	0.131
		row_to_text_list	True	-2.317	0.306	0.511	0.306	0.400	0.516	0.991	0.104
			False	-2.477	0.419	0.671	0.419	0.500	0.527	0.994	0.136
	most_cosine_similar	row_to_text_str	True	-2.044	0.349	0.237	0.349	0.800	0.422	0.882	0.137
			False	-2.231	0.336	0.424	0.336	0.500	0.527	0.937	0.140
		row_to_text_list	True	-2.392	0.353	0.585	0.353	0.400	0.516	0.989	0.138
			False	-2.545	0.463	0.738	0.463	0.300	0.483	1.029	0.148
	least_cosine_similar	row_to_text_str	True	-2.416	0.232	0.610	0.232	0.500	0.527	1.068	0.189
			False	-2.435	0.287	0.628	0.287	0.200	0.422	1.053	0.096
		row_to_text_list	True	-2.365	0.325	0.559	0.325	0.600	0.516	1.069	0.197
			False	-2.651	0.342	0.844	0.342	0.200	0.422	1.054	0.147
	cosine_diverse	row_to_text_str	True	-2.206	0.402	0.399	0.402	0.600	0.516	0.981	0.110
			False	-2.787	0.385	0.980	0.385	0.100	0.316	1.047	0.120
		row_to_text_list	True	-2.430	0.481	0.624	0.481	0.200	0.422	0.941	0.170
			False	-2.611	0.355	0.805	0.355	0.200	0.422	1.038	0.143

Table 5: **Study 1 – POLIDEOLOGY.** Simulation results for "Llama-3.1-8B-Instruct" with various retriever configurations for the IV POLIDEOLOGY.

few-shot examples	retrieval strategy	row conversion strategy	stratification	coef mean	coef std	bias mean	bias std	coverage mean	coverage std	width mean	width std
100	random_selection_static_context	row_to_text_str	True	-0.068	0.121	0.030	0.121	0.900	0.316	0.466	0.033
			False	-0.035	0.177	-0.004	0.177	1.000	0.000	0.635	0.043
		row_to_text_list	True	-0.067	0.111	0.028	0.111	0.900	0.316	0.464	0.033
			False	-0.021	0.154	-0.018	0.154	1.000	0.000	0.633	0.076
	random_selection	row_to_text_str	True	-0.054	0.117	0.015	0.117	0.900	0.316	0.466	0.029
			False	-0.024	0.145	-0.015	0.145	1.000	0.000	0.622	0.071
		row_to_text_list	True	-0.059	0.102	0.020	0.102	0.900	0.316	0.463	0.029
			False	-0.018	0.149	-0.021	0.149	1.000	0.000	0.654	0.076
	most_cosine_similar	row_to_text_str	True	-0.044	0.134	0.005	0.134	0.900	0.316	0.542	0.042
			False	0.010	0.121	-0.048	0.121	1.000	0.000	0.617	0.063
		row_to_text_list	True	-0.022	0.106	-0.017	0.106	1.000	0.000	0.489	0.029
			False	-0.019	0.150	-0.020	0.150	1.000	0.000	0.642	0.089
	least_cosine_similar	row_to_text_str	True	-0.045	0.102	0.006	0.102	1.000	0.000	0.490	0.033
			False	-0.001	0.126	-0.038	0.126	1.000	0.000	0.604	0.043
		row_to_text_list	True	-0.032	0.112	-0.007	0.112	0.900	0.316	0.457	0.023
			False	0.006	0.107	-0.045	0.107	1.000	0.000	0.620	0.046
	cosine_diverse	row_to_text_str	True	-0.059	0.100	0.020	0.100	0.900	0.316	0.453	0.025
			False	-0.053	0.135	0.015	0.135	1.000	0.000	0.642	0.050
		row_to_text_list	True	-0.067	0.110	0.028	0.110	0.900	0.316	0.467	0.035
			False	-0.030	0.159	-0.009	0.159	1.000	0.000	0.627	0.044
10	random_selection_static_context	row_to_text_str	True	-0.052	0.197	0.013	0.197	0.900	0.316	0.579	0.068
			False	-0.048	0.177	0.009	0.177	1.000	0.000	0.597	0.041
		row_to_text_list	True	-0.053	0.181	0.015	0.181	0.900	0.316	0.576	0.066
			False	-0.047	0.180	0.008	0.180	1.000	0.000	0.582	0.036
	random_selection	row_to_text_str	True	-0.056	0.120	0.017	0.120	1.000	0.000	0.594	0.057
			False	-0.031	0.133	-0.008	0.133	1.000	0.000	0.594	0.041
		row_to_text_list	True	-0.046	0.121	0.008	0.121	1.000	0.000	0.590	0.054
			False	-0.049	0.153	0.010	0.153	1.000	0.000	0.606	0.061
	most_cosine_similar	row_to_text_str	True	-0.127	0.119	0.088	0.119	0.900	0.316	0.559	0.048
			False	-0.046	0.092	0.008	0.092	1.000	0.000	0.605	0.049
		row_to_text_list	True	-0.114	0.141	0.075	0.141	1.000	0.000	0.590	0.071
			False	-0.044	0.141	0.005	0.141	1.000	0.000	0.596	0.044
	least_cosine_similar	row_to_text_str	True	-0.062	0.125	0.024	0.125	0.900	0.316	0.577	0.039
			False	-0.042	0.114	0.003	0.114	1.000	0.000	0.587	0.100
		row_to_text_list	True	0.005	0.111	-0.044	0.111	1.000	0.000	0.568	0.053
			False	0.043	0.134	-0.082	0.134	0.900	0.316	0.600	0.038
	cosine_diverse	row_to_text_str	True	-0.070	0.114	0.031	0.114	1.000	0.000	0.583	0.073
			False	0.026	0.168	-0.065	0.168	0.900	0.316	0.583	0.033
		row_to_text_list	True	-0.074	0.181	0.036	0.181	0.900	0.316	0.576	0.051
			False	-0.042	0.157	0.003	0.157	1.000	0.000	0.588	0.033

Table 6: **Study 1 – RELIGATTEND.** Simulation results for ”Llama-3.1-8B-Instruct” with various retriever configurations for the IV RELIGATTEND.