Supplementary Information (Dataverse-only) for "Improving Probabilistic Models in Text Classification via Active Learning"*

Mitchell Bosley[†‡] Saki Kuzushima[†§] Ted Enamorado[¶]
Yuki Shiraito[‖]

First draft: September 10, 2020
Final submission: April 22, 2024

In this supplementary information (Dataverse-only), we provide further details on: 1) The validation performance of *activeText* , 2) The reanalyses using *activeText* of Gohdes (2020) and Park et al. (2020), and 3) Our simulation studies.

H Supplemental Results for "Validation Performance"

We present additional results supplementing those presented in the Section "Validation Performance."

*We thank Ken Benoit, Yaoyao Dai, Chris Fariss, Yusaku Horiuchi, Kosuke Imai, Walter Mebane, Daichi Mochihashi, Kevin Quinn, Luwei Ying, audiences at the 2020 Annual Meeting of the American Political Science Association, the 2021 Annual Meeting of the Midwest Political Science Association, the 11th Annual Conference on New Directions in Analyzing Text as Data, the 2022 Summer Meeting of the Japanese Society for Quantitative Political Science, and the 40th Annual Summer Meeting of the Society for Political Methodology, and seminar participants at the University of Michigan and members of the Junior Faculty Workshop at Washington University in St. Louis for useful comments and suggestions. We also appreciate detailed and constructive comments from four anonymous reviewers of the journal. Finally, we are extremely grateful to the editor, Michelle Dion, for guiding us through the rigorous review process of *APSR*.

†These authors have contributed equally to this work.

‡Ph.D. Candidate, Department of Political Science, University of Michigan. Email: mcbosley@umich.edu. ORCID: 0000-0002-9172-966X.

§Ph.D. Candidate, Department of Political Science, University of Michigan. Email: skuzushi@umich.edu. ORCID: 0000-0003-3014-5203.

¶Assistant Professor, Department of Political Science, Washington University in St. Louis. Siegle Hall, 244. One Brookings Dr. St Louis, MO 63130-4899. Phone: 314-935-5810, Email: ted@wustl.edu, URL: www.tedenamorado.com. ORCID: 0000-0002-2022-7646.

‖Assistant Professor, Department of Political Science, University of Michigan. Center for Political Studies, 4259 Institute for Social Research, 426 Thompson Street, Ann Arbor, MI 48104-2321. Phone: 734-615-5165, Email: shiraito@umich.edu, URL: shiraito.github.io. ORCID: 0000-0003-0264-1138.

H.1 Classification Performance

To complement the results presented in Figure 2 in the main text, Table H.1 presents the results (across datasets) of fitting our model at the initial (iteration 0) and last active step (iteration 30). It is clear from the table that the improvements *activeText* brings in terms of the F1-score, precision, and recall. Furthermore, after labeling 600 documents (20 per iteration), uncertainty sampling outperforms random sampling across evaluation metrics, which empirically validates the promise of active learning in terms of text classification.

Table H.1: **Classification Performance: Uncertainty vs Random Sampling with** $\lambda = 0.001$

Dataset	Active Step	Uncertainty Sampling			Random Sampling		
		Precision	Recall	F1-score	Precision	Recall	F1-score
`Wikipedia`	0	0.71	0.13	0.22	0.71	0.13	0.22
	30	0.71	0.54	0.61	0.45	0.56	0.50
`BBC`	0	0.33	0.86	0.48	0.33	0.86	0.48
	30	0.92	0.96	0.94	0.92	0.94	0.93
`Supreme Court`	0	0.46	0.98	0.63	0.46	0.98	0.63
	30	0.85	0.91	0.88	0.75	0.96	0.84
`Human Rights`	0	0.61	0.01	0.02	0.61	0.01	0.02
	30	0.53	0.42	0.47	0.46	0.44	0.45

Similarly, and as noted in the main text, our results appear to be not too sensitive to the selection of the weighting parameter λ, provided that its value remains small. Figures H.1 and H.2 confirm this finding. Figure H.1, shows that if compared to passive learning, after 30 active steps, the performance of *activeText* is better in terms of F1-score when $\lambda = 0.001$ if compared to $\lambda = 0.01$. Figure H.2 demonstrates that when λ is small (for instance, when $\lambda = 0$ representing a supervised model, or $\lambda = 0.001$), the out-of-sample performance of *activeText* , as indicated by the F1-score, is notably superior compared to cases where λ is large (for example, when $\lambda = 0.5$ or $\lambda = 1$, representing models where the labeled and unlabeled data carry equal weight).

H.2 Visual Demonstration of Updating the Word-Class Matrix for the Human Rights Corpus

In Subsection "Benefits of Keyword Upweighting," we highlight a possible explanation for the initial poor performance of *activeText* with the Human Rights Allegation dataset: the length of the documents. Unlike other corpora, each document in the human rights corpus comprises only one sentence on average. Consequently, the amount of information available

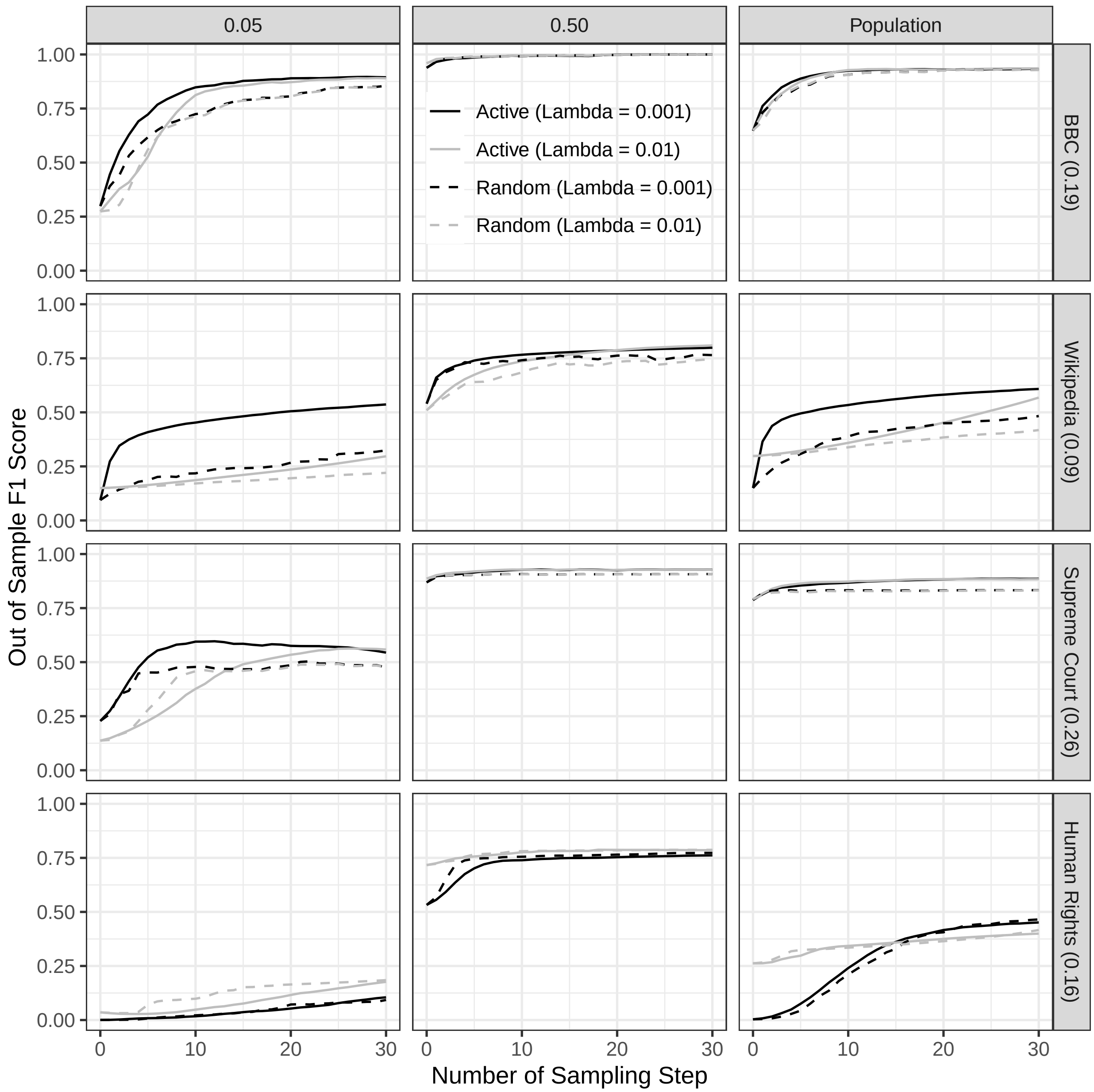

Figure H.1: **Classification Results with 2 Classes and $\lambda = 0.01$ vs $\lambda = 0.001$.** The darker lines show the results with $\lambda = 0.001$ and the lighter lines show $\lambda = 0.01$. The columns correspond to various proportion of positive labels in the corpus. The y-axis indicates the out-of-sample F1 score and the x-axis show the number of sampling steps. The smaller the value of λ the better the performance of our model.

for the model to learn from labeled documents is comparatively limited. In such scenarios, supplementing document labels with keywords can enhance classification accuracy. Our findings in Section Benefits of Keyword Upweighting indicate that incorporating keywords indeed enhances performance, particularly in cases where there is a significant imbalance in

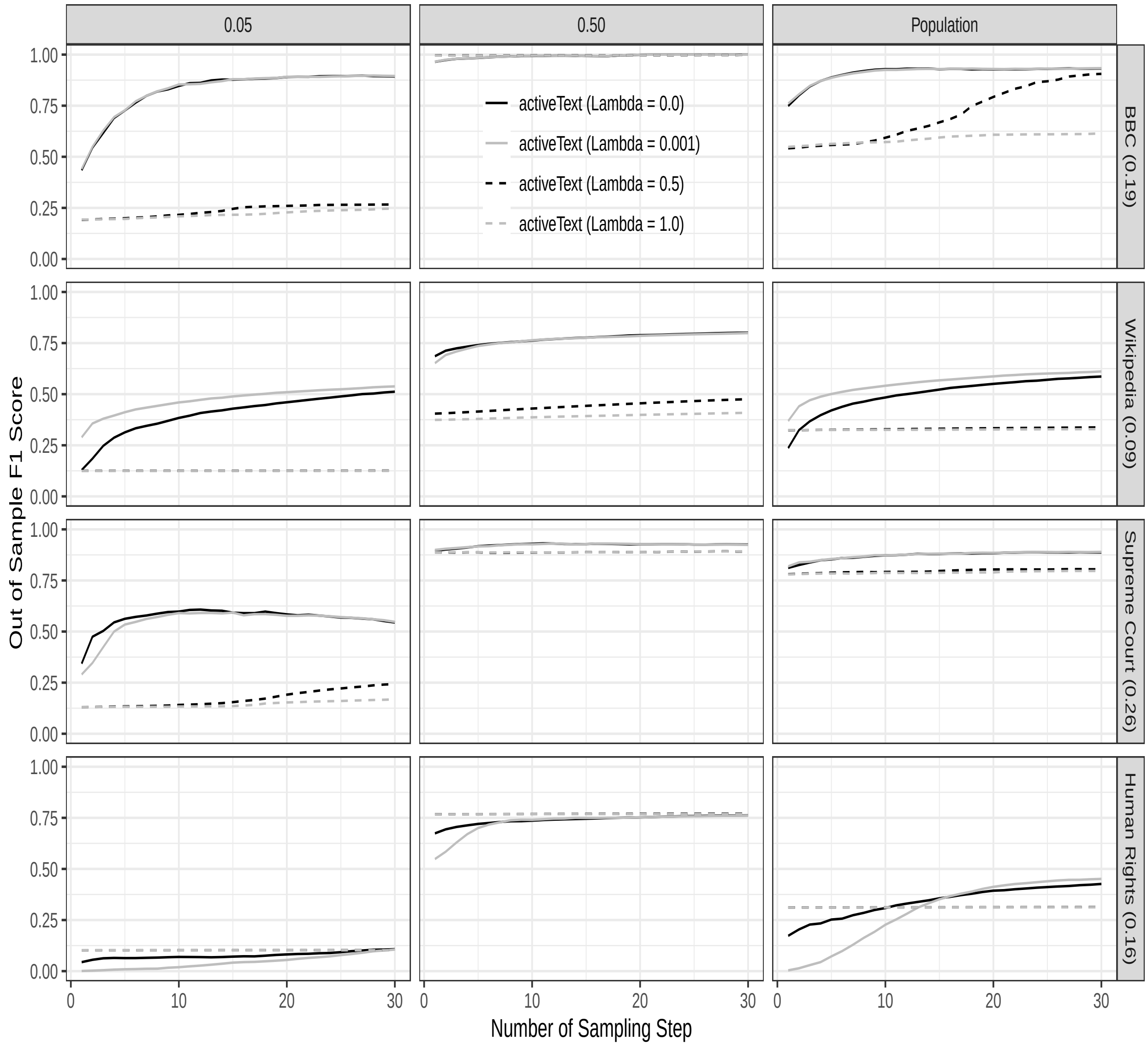

Figure H.2: **Classification Results with 2 Classes and $\lambda \in \{0, 0.001, 0.5, 1\}$.**
The columns correspond to various proportion of positive labels in the corpus. The y-axis indicates the out-of-sample F1 score and the x-axis show the number of sampling steps. The smaller the value of λ the better the performance of our model.

document distribution across classes. In this SI, we visually illustrate why is this the case.

Figure H.3 visually demonstrates how the word-class matrix $\boldsymbol{\eta}$ is updated across iterations with and without the inclusion of keywords. In the generative model for *activeText* , the importance of words for classification hinges on two factors. The first is the ratio of the probability that a word appears in a document with a positive vs. negative class, and the second is the word frequency. As these two values increase, the importance of a word increases. In Figure H.3, the former is presented on the x-axis and the latter on the y-axis.

Specifically, the x-axis shows the log of η_{v1}/η_{v0}, where η_{v1} corresponds the probability of observing the word v in a document with a positive class and η_{v0} for a document with a negative class, and the y-axis is the log of word frequency. This means that words towards the upper right corner of the figure are more important for the classification. Therefore, this visualization allows users to interpret which words are more important to the classification.

This visualization also illustrates how the keywords are used to update the word-class matrix. By shifting the value of $\boldsymbol{\eta}$ of keywords, which typically appear on the right side of the figures, we can accelerate the estimation of $\boldsymbol{\eta}$ and improve the classification performance. A subset of the keywords supplied is labeled and highlighted by black dots.

H.3 Main Results including SVM Comparisons

Previous iterations of our research incorporated comparisons utilizing a Support Vector Machine (SVM) classifier with Active Learning, as detailed by Miller et al. (2020). In order to preserve the memory of those findings, Figure H.4 complements Figure 2 in the main text by presenting additional comparisons involving both passive and active learning algorithms within the SVM framework (Miller et al., 2020). Note that we use the *margin sampling* variation of their model, which they showed performed the best in their analysis. In order to facilitate a fair comparison with our models, we evaluate their model with our Quanteda-based DFMs, rather than the SciKit Learn-based matrices in the original analysis. As a result, the cross-validation-of-DFMs feature in their original analysis is omitted here.

For *activeText* , we set the λ parameter to be 0.001.[1] In all specifications, we use *entropy sampling* to select the documents that will be labeled in each active learning iteration, arranging the unlabeled documents in descending order in terms of Shannon entropy, then selecting the top n documents. Additionally, the reported results are the average of 100 Monte Carlo iterations for each model. In each Monte Carlo iteration, the model is randomly initialized with 20 documents and in each active iteration (see Algorithm 1) 20 additional documents are labeled.[2]

As Figure H.4 shows, our overarching conclusions persist: in comparison to both Bidirectional Encoder Representations from Transformers (BERT) and SVM models, *activeText* consistently exhibits superior performance during the initial phases of the evaluation process. The sole exception to this trend occurs with the Human Rights dataset, where BERT demonstrates superior performance over *activeText* .

[1]We decided to use 0.001 because we found that this value results in a good performance across the datasets. To the extent that λ is kept small i.e., less than 0.01, our main findings remain unchanged (see Figure H.1).

[2]In order to ensure a fair comparison with the active SVM model, the *activeText* and SVM models are initialized with the same random documents for each Monte Carlo iteration.

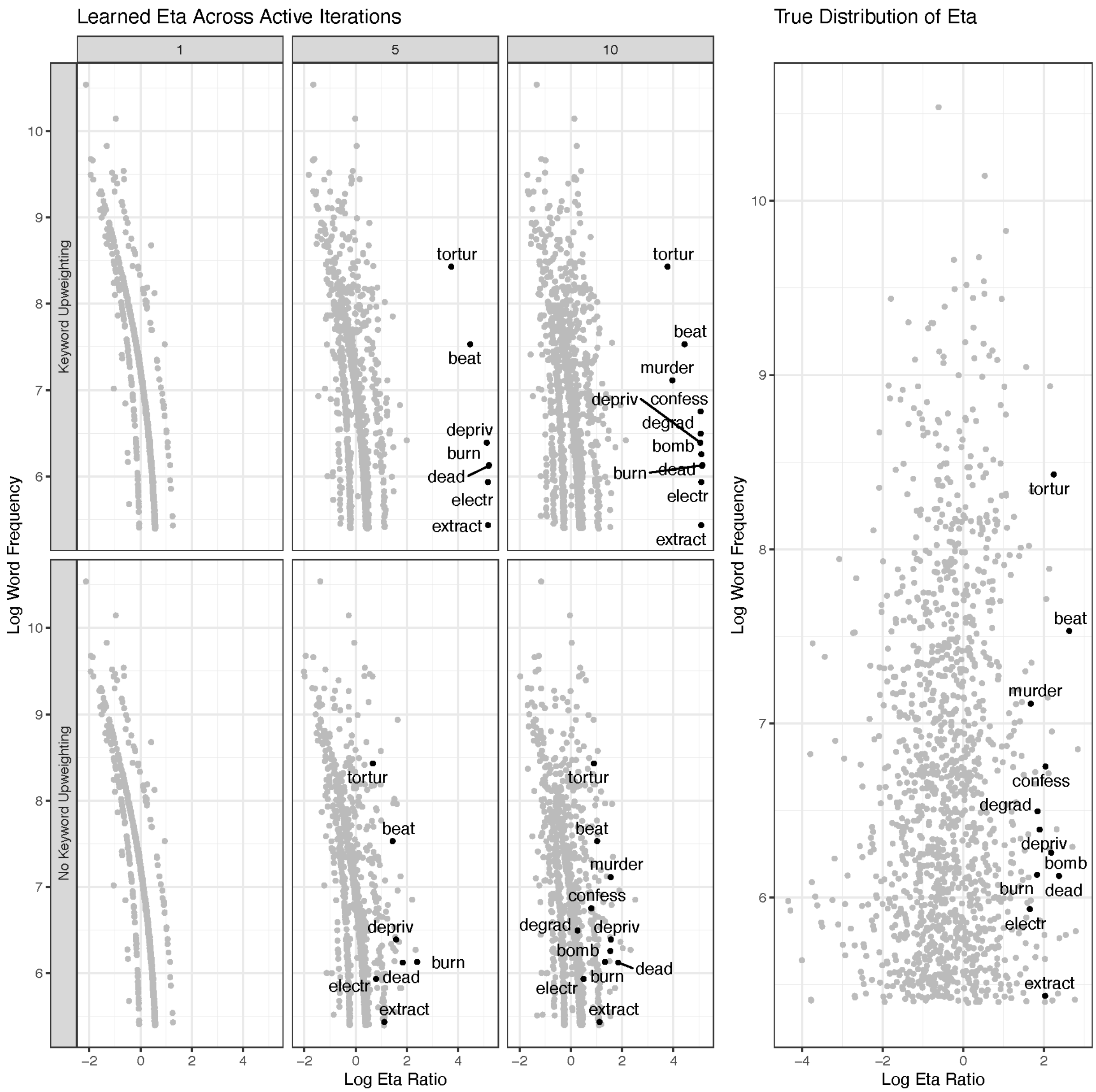

Figure H.3: **Update of the Word-class Matrix ($\boldsymbol{\eta}$) with and without Keywords** The right figure shows the distribution of the estimated word-class matrix $\boldsymbol{\eta}$ with fully labeled corpus. The left figure shows how $\boldsymbol{\eta}$ is updated across active iterations. The top row shows the updating process of $\boldsymbol{\eta}$ with keywords and the bottom row without. The first column shows the initial values of $\boldsymbol{\eta}$, the middle column shows $\boldsymbol{\eta}$ after 5 iterations (100 labels), and the left columns after 10 iterations (200 labels). The x-axis shows the log ratio of η_{v1} vs η_{v0}, where $K = 2$ is linked to the positive class. If this value is high, a word v is more strongly associated with positive labels. The y-axis is the log of word frequency. A word with high word frequency has more influence in shifting the label probability. A subset of keywords are labeled and highlighted by black dots. Keywords scheme accelerates the learning process of $\boldsymbol{\eta}$ by upweighting the value of corresponding η_{vk} in the positive direction.

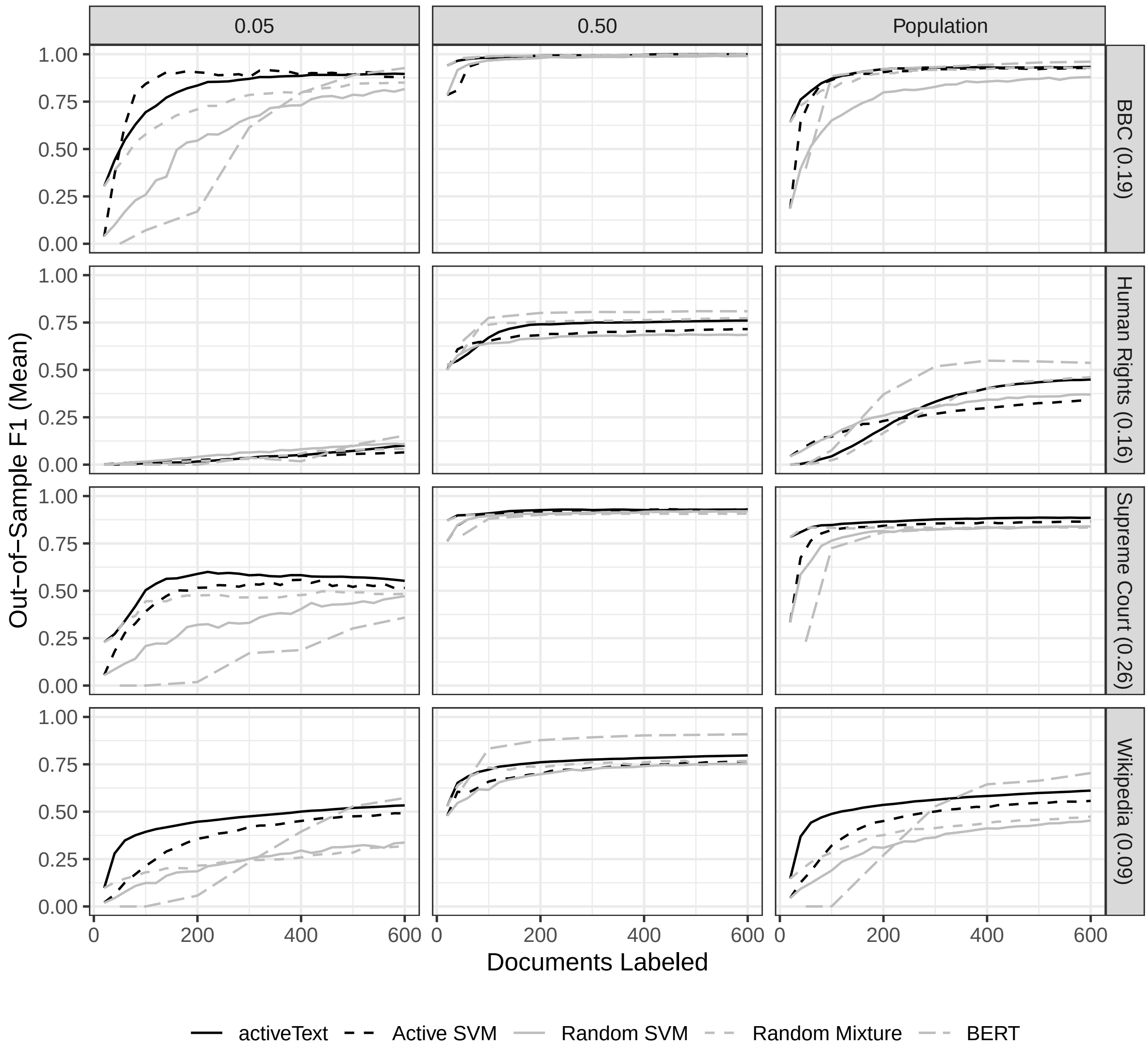

Figure H.4: **Replication of F1 performance from Figure 2 with random and active SVM**

The rows correspond to different datasets and the columns correspond to various proportion with positive label documents in the corpus. The y-axis indicates the out-of-sample F1 score and the x-axis shows the number of sampling steps. 20 documents are labeled at each sampling step. The colors correspond to different classifiers. The line type shows different sampling schemes: the solid lines are for the active sampling and the dashed line are for the random sampling. Active learning often performs better than passive learning, especially the proportion of positive labels are small. Moreover *activeText* performs favorably if compared to state-of-the-art methods such as BERT.

I Supplemental Results for the Reanalysis of "Repression Technology: Internet Accessibility and State Violence" (Gohdes, 2020)

In this section, we provide additional findings to complement those discussed in Section "Internet Accessibility and State Violence (Gohdes, 2020)". Specifically, we begin by comparing the accuracy of predictions made by Gohdes (2020) with those made by *activeText* , focusing on the construction of the proportion of biweekly government-targeted killings, which is a key variable in Gohdes (2020). Additionally, we present tables that showcase the complete regression results as reported in Gohdes (2020), vis-à-vis with regression specifications utilizing the proportion of biweekly government-targeted killings from *activeText* .

I.1 Comparison of the Predictions Between *activeText* and `XGboost` predictions

Table I.1 shows a crosstable between the prediction based on *activeText* and the predictions from Gohdes (2020) obtained using `XGboost`.[3] As the Table shows, most observations fall in the diagonal cells of the matrix, and the correlation between the two predictions is high (0.93). One difference is that *activeText* classifies more documents to target killings compared to the original predictions. Note that neither prediction claims to be the ground truth, both are the results of different classifiers.

In Section "Internet Accessibility and State Violence (Gohdes, 2020)," the method used by the author involves utilizing predictions generated by `XGboost`. These predictions are then used to determine the proportion of biweekly government-targeted killings for each of the 14 Syrian governorates, which are the second largest administrative divisions in Syria. Gohdes (2020) collapses these predictions at the governorate-biweekly level. This proportion serves as the primary outcome measure in the analysis. Figure I.1 supports the results presented in Table I.1. Essentially, when comparing the proportion of biweekly government-targeted killings constructed via `XGboost` with the same construct in Gohdes (2020), most data points align closely with the 45-degree line, indicating a high level of agreement.

[3] `XGboost` is a scalable, distributed gradient-boosted decision tree often used for classification and other prediction tasks.

		Original		
		untargeted	targeted	non-government
	untargeted	50327	411	135
activeText	targeted	1630	10044	31
	non-government	382	34	2280

Table I.1: Cross Table between *activeText* and `XGboost` predictions

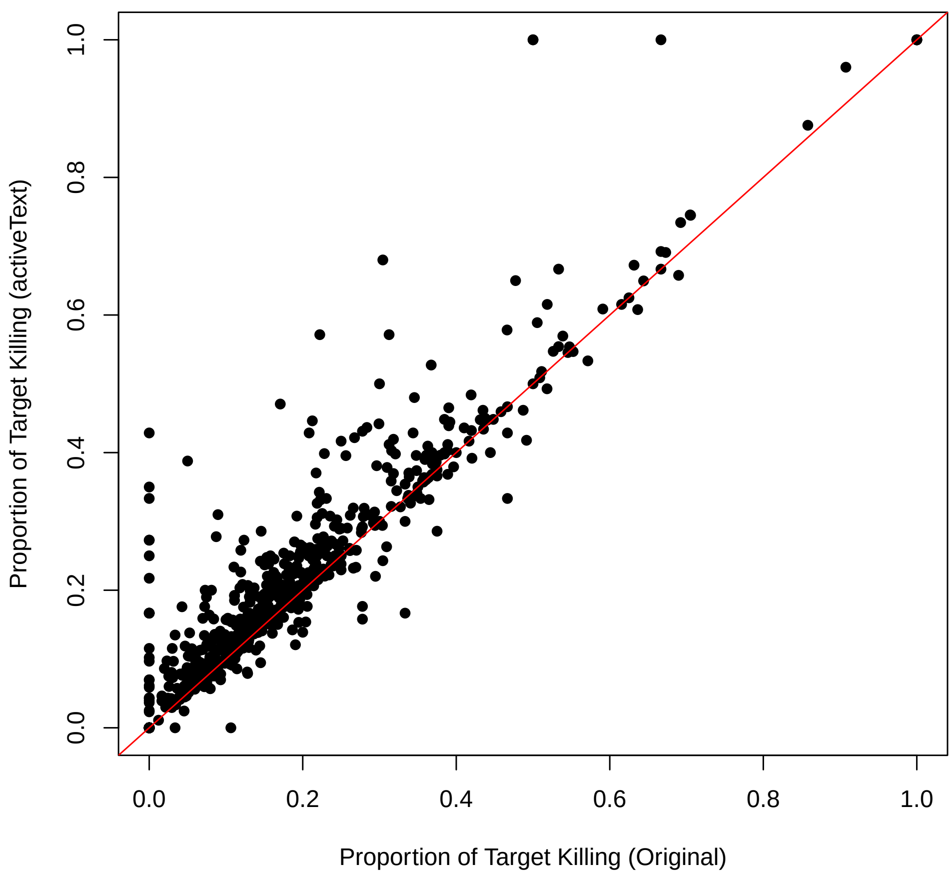

Figure I.1: **Scatter plot of the dependent variable between the one constructed by *activeText* vs. the original constructed by Gohdes (2020)**
The author performs a binomial logit regression where the dependent variable is the ratio of the number of targeted killings to the total number of government killings. We compare the dependent variable used in the original paper vs. the one we constructed using *activeText* . The 45-degree line (in red) corresponds to equality between measures. We can see that most observations lie around the 45-degree line while there are some values in the upper triangle. This suggests that *activeText* yields a similar dependent variable to the original one, while there may be some overestimations of the proportion of target killing with *activeText* .

I.2 Regression Results in Gohdes (2020)

Table I.2 presents the original regression table reported in Gohdes (2020), while Table I.3 is a replication of the original table using *activeText* instead. In both tables, the coefficient on associated to the main predictor, Internet access, are positive and statistically significant, which match the author's substantive conclusion. However, when comparing both tables, there is a discrepancy between the magnitude of the coefficients associated with IS control (Islamic State control), and the interaction between IS control and Internet access (3G). We believe that this is because the number of observations in the IS control is small (51) and there is almost no variation of the Internet access variable within the observations with IS control, as shown in Figure I.2.

	I	II	III	IV	V	VI	VII
Intercept	-2.340^{***}	-2.500^{***}	-0.899^{*}	-0.410	-0.019	-1.308	-3.013^{**}
	(0.205)	(0.267)	(0.403)	(0.521)	(0.357)	(1.057)	(1.103)
Internet access (3G)	0.224^{*}	0.231^{*}	0.200^{*}	0.205^{*}	0.265^{*}	0.313^{**}	0.909^{***}
	(0.095)	(0.094)	(0.085)	(0.087)	(0.113)	(0.116)	(0.124)
% Govt control							0.016^{***}
							(0.004)
Internet (3G) * % Govt control							-0.014^{***}
							(0.001)
Govt control	0.774^{*}	0.803^{**}	1.167^{***}	1.180^{***}	0.080	0.856^{**}	0.811^{***}
	(0.332)	(0.272)	(0.284)	(0.288)	(0.344)	(0.313)	(0.237)
IS control	2.027^{***}	1.644^{***}	1.045^{*}	-0.324	0.432	0.787	-0.663^{**}
	(0.435)	(0.462)	(0.421)	(0.209)	(0.414)	(0.418)	(0.221)
Kurd control	0.386	-0.243	-0.506	-1.331	-0.402	0.033	-0.616
	(0.594)	(0.843)	(0.760)	(1.134)	(0.745)	(0.802)	(0.432)
Opp control	1.160^{***}	1.252^{***}	0.727^{*}	0.759^{*}	-0.700^{*}	-0.281	-0.176
	(0.298)	(0.317)	(0.293)	(0.296)	(0.283)	(0.342)	(0.164)
Internet (3G) * Govt control	-0.163	-0.182	-0.327^{**}	-0.324^{**}	-0.104	-0.358^{**}	
	(0.132)	(0.117)	(0.119)	(0.122)	(0.133)	(0.120)	
Internet (3G) * IS control	-1.798^{***}	-1.525^{***}	-1.377^{***}		-1.391^{***}	-1.336^{***}	
	(0.220)	(0.281)	(0.251)		(0.264)	(0.261)	
Internet (3G) * Kurd control	-0.133	0.336	0.093	0.895	-0.052	-0.202	
	(0.444)	(0.649)	(0.569)	(0.936)	(0.553)	(0.527)	
Internet (3G) * Opp. control	-0.605^{***}	-0.722^{***}	-0.511^{**}	-0.533^{***}	0.316^{*}	0.286	
	(0.159)	(0.173)	(0.157)	(0.158)	(0.151)	(0.186)	
# Killings (log)			-0.273^{***}	-0.271^{***}	-0.354^{***}	-0.412^{***}	-0.584^{***}
			(0.054)	(0.055)	(0.051)	(0.072)	(0.074)
Govt gains				0.643			
				(0.385)			
Govt losses				0.632			
				(0.413)			
Christian					0.068	0.345^{**}	0.398^{***}
					(0.111)	(0.116)	(0.110)
Alawi					1.479^{**}	-1.167^{***}	-0.812^{***}
					(0.522)	(0.177)	(0.176)
Druze					-0.634^{***}	-0.302	0.135
					(0.191)	(0.191)	(0.190)
Kurd					-0.659^{***}	-0.542^{*}	-0.580^{**}
					(0.194)	(0.237)	(0.212)
Internet (3G) * Alawi					-0.909^{***}		
					(0.163)		
Pop (log)						0.196	0.408^{**}
						(0.149)	(0.150)
Unempl. (%)						-0.016	-0.002
						(0.012)	(0.012)
AIC	11956.847	9993.704	9665.749	9495.591	7671.979	7873.915	7327.796
BIC	12001.524	10239.427	9915.941	9744.552	7944.509	8150.913	7595.858
Log Likelihood	−5968.424	−4941.852	−4776.875	−4691.796	−3774.990	−3874.958	−3603.898
Deviance	9519.651	7466.508	7136.554	7026.891	5132.784	5332.720	4790.601
Num. obs.	640	640	640	626	640	640	640

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Reference category: Contested control. Governorate-clustered SEs.

Table I.2: Table 1 in Gohdes 2020: Original table

	I	II	III	IV	V	VI	VII
Intercept	-2.196^{***}	-2.428^{***}	-0.795^{*}	-0.351	-0.037	-1.141	-2.695^{*}
	(0.197)	(0.242)	(0.390)	(0.490)	(0.348)	(1.229)	(1.227)
Internet access (3G)	0.277^{**}	0.282^{***}	0.242^{**}	0.250^{**}	0.342^{***}	0.369^{***}	0.853^{***}
	(0.091)	(0.081)	(0.075)	(0.077)	(0.103)	(0.107)	(0.118)
% Govt control							0.015^{***}
							(0.004)
Internet (3G) * % Govt control							-0.013^{***}
							(0.001)
Govt control	0.625^{*}	0.672^{**}	1.048^{***}	1.058^{***}	0.151	0.843^{**}	0.559^{*}
	(0.319)	(0.255)	(0.269)	(0.273)	(0.358)	(0.300)	(0.249)
IS control	15.157^{***}	15.688^{***}	15.072^{***}	-0.275	14.551^{***}	14.877^{***}	-0.600^{**}
	(1.123)	(1.148)	(1.136)	(0.200)	(1.132)	(1.134)	(0.209)
Kurd control	0.795	0.099	-0.227	-0.440	-0.157	0.334	-0.369
	(0.516)	(0.729)	(0.671)	(1.119)	(0.677)	(0.744)	(0.405)
Opp control	0.978^{***}	1.134^{***}	0.594^{*}	0.634^{*}	-0.606^{*}	-0.197	-0.278
	(0.294)	(0.304)	(0.284)	(0.289)	(0.270)	(0.322)	(0.155)
Internet (3G) * Govt control	-0.169	-0.190	-0.334^{**}	-0.335^{**}	-0.183	-0.408^{***}	
	(0.126)	(0.103)	(0.108)	(0.111)	(0.131)	(0.111)	
Internet (3G) * IS control	-14.829^{***}	-15.506^{***}	-15.351^{***}		-15.392^{***}	-15.330^{***}	
	(1.080)	(1.096)	(1.090)		(1.091)	(1.091)	
Internet (3G) * Kurd control	-0.400	0.138	-0.080	0.134	-0.240	-0.366	
	(0.324)	(0.514)	(0.463)	(0.940)	(0.473)	(0.460)	
Internet (3G) * Opp. control	-0.542^{***}	-0.688^{***}	-0.468^{**}	-0.497^{**}	0.181	0.149	
	(0.159)	(0.164)	(0.150)	(0.152)	(0.145)	(0.176)	
# Killings (log)			-0.278^{***}	-0.274^{***}	-0.356^{***}	-0.415^{***}	-0.567^{***}
			(0.053)	(0.054)	(0.051)	(0.071)	(0.073)
Govt gains				0.512			
				(0.349)			
Govt losses				0.730^{*}			
				(0.334)			
Christian					0.092	0.352^{**}	0.369^{***}
					(0.115)	(0.113)	(0.105)
Alawi					1.329^{*}	-0.928^{***}	-0.585^{***}
					(0.528)	(0.167)	(0.168)
Druze					-0.628^{**}	-0.310	0.063
					(0.196)	(0.197)	(0.209)
Kurd					-0.565^{**}	-0.502^{*}	-0.615^{**}
					(0.204)	(0.227)	(0.207)
Internet (3G) * Alawi					-0.782^{***}		
					(0.164)		
Pop (log)						0.185	0.391^{*}
						(0.167)	(0.168)
Unempl. (%)						-0.019	-0.007
						(0.012)	(0.012)
AIC	12050.644	10116.531	9739.975	9570.556	8038.596	8197.433	7735.527
BIC	12095.321	10362.255	9990.166	9819.517	8311.125	8474.431	8003.589
Log Likelihood	-6015.322	-5003.266	-4813.988	-4729.278	-3958.298	-4036.717	-3807.763
Deviance	9500.059	7475.946	7097.391	6986.658	5386.011	5542.849	5084.942
Num. obs.	640	640	640	626	640	640	640

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Reference category: Contested control. Governorate-clustered SEs.

Table I.3: Table 1 in Gohdes 2020: Reanalysis with *activeText*

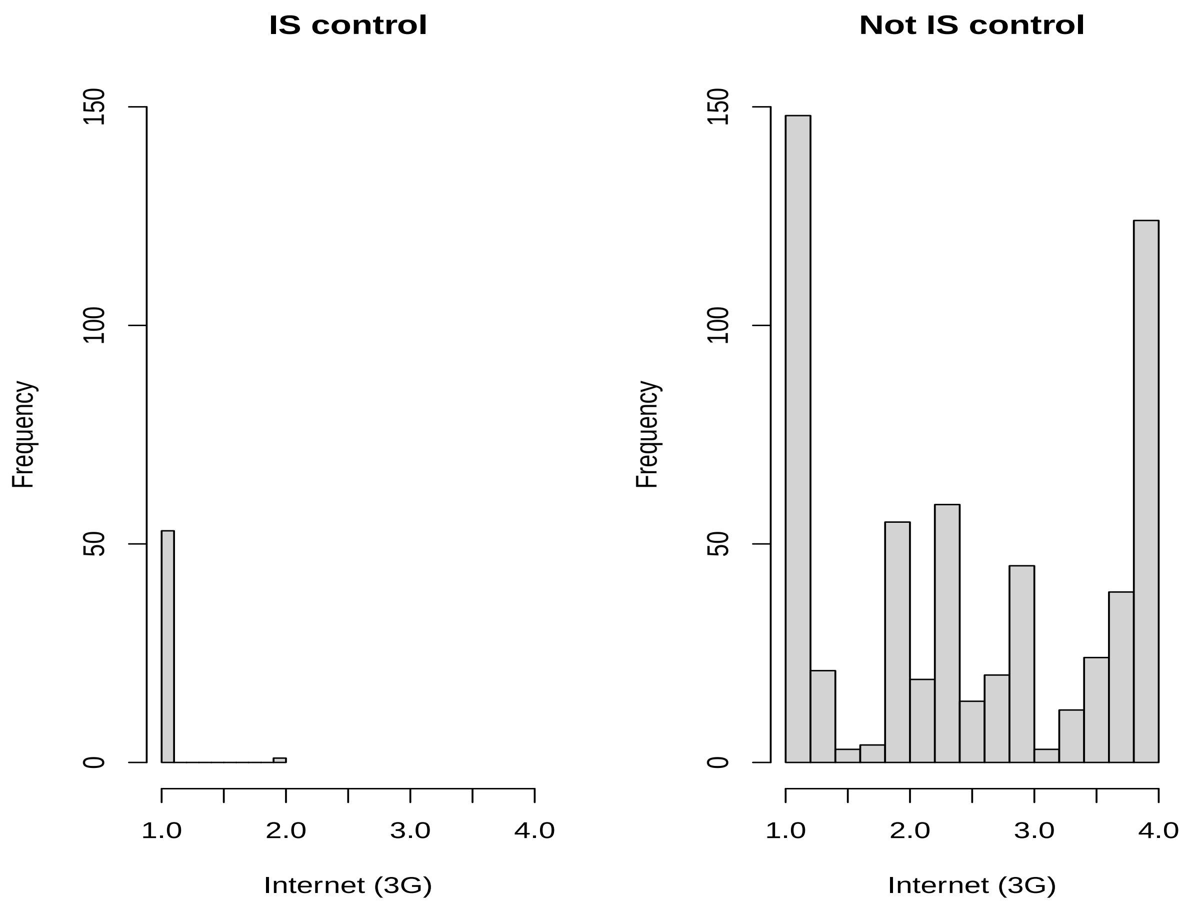

Figure I.2: **Histogram of the Internet (3G) variable by the presence of Islamic State (IS) control in the original data**
The left histogram is the distribution of the Internet (3G) variable for the observation under IS control, and the right one is not under IS control. The number of observations with IS control is only 51 out of the total observation of 640. In addition, among those with IS control, all observations except one takes the same value for the Internet access variable. This suggests that the regression coefficient on the interaction of IS control and Internet access can be highly unstable.

J Supplemental Results for the Reanalysis of "Human Rights are Increasingly Plural" (Park et al., 2020)

In this section, we present additional results mentioned in the main text about our reanalysis of Park et al. (2020). As mentioned in the main text, we follow the same pre-processing approach as Park et al. (2020) when analyzing their dataset of Country Reports on Human Rights Practices spanning from 1977 to 2016, sourced from the US Department of State. This involves utilizing their set of 4000 human rights reports, which have been manually labeled as positive, negative, or neutral (with counts of 1182 positive, 1743 negative, and 1075 neutral reports), and employing the same document-feature matrices comprising 30,000 features, including both unigrams and bigrams.

We stopped our active labeling process after the 25th iteration of our algorithm. As we discuss in Section "The Method", this decision was based on observing that the out-of-sample F1 score, derived from an 80/20 training/test split, did not increase by more than 0.01 units beyond this point (refer to Figure J.1).

Figure 6 in the main text illustrates that by labeling only 500 documents with *activeText* , instead of the 4000 documents labeled by Park et al. (2020) for their SVM classifier, we reach a similar substantive conclusion: that the average sentiment of human rights reports has remained relatively stable and close to neutral over time. Additionally, in Figure J.2, we demonstrate that this finding is not influenced by our stopping rule, and it remains robust even when labeling more documents, such as 1000, 1500, and 2000, instead of just 500.

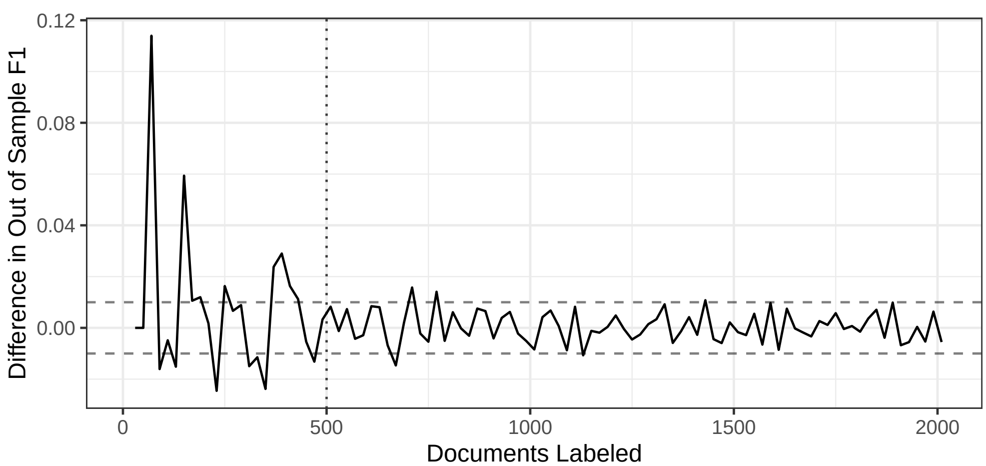

Figure J.1: **Using the Difference in Out of Sample F1 Score to Decide a Stopping Point.**
As shown in the figure, the out-of-sample F1 score does not increase by more than 0.01 units beyond labeling 500 documents.

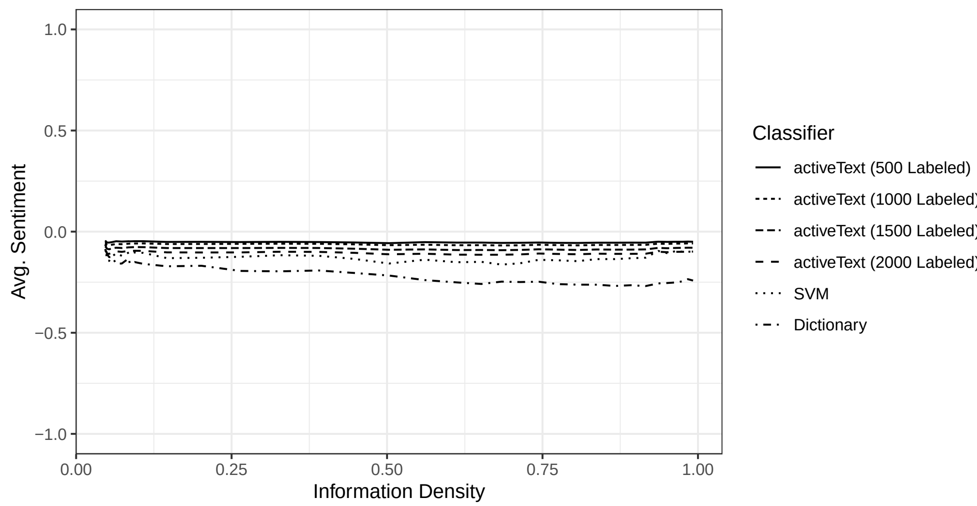

Figure J.2: **Replication of Figure 1 in Park et al. (2020): The Relationship Between Information Density and Average Sentiment Score Across Different Settings for the Total Number of Labeled Documents.**
The figure shows that, when labeling more documents (such as 1000, 1500, and 2000, instead of just 500), the results presented in the main text are almost identical and the substantive conclusions unchanged.

K Supplementary Appendix for the Simulation Studies

This section provides a comprehensive presentation of the simulation results. The tables display the average classification performance achieved in each simulation configuration. We assess the classification performance using four metrics: accuracy, precision, recall, and F1 score. We generate 100 datasets for each simulation setup and report the average performance obtained across these datasets.

Two key findings consistently emerge across all simulation setups. First active learning demonstrates superior performance compared to passive learning (a method that, like LURE, does not suffer from in-sample bias in training), especially during the initial stages of hand-labeling. This implies that the benefits of active learning are most pronounced when a small fraction of documents can be manually labeled, a scenario commonly encountered in practical situations. Once a substantial portion of documents has been hand-labeled, passive learning achieves comparable performance to active learning. This observation is unsurprising since only a limited number of documents remain for labeling.

Second, active learning outperforms passive learning in datasets with imbalanced proportions between classes. In cases where the proportion of classes is balanced, random sampling performs as well as active learning.

Number of documents	Vocabulary size	Avg. number of words per doc.	β	Table
1000	1000	10	0.1	Table K.2
			0.5	Table K.3
			0.9	Table K.4
		50	0.1	Table K.5
			0.5	Table K.6
			0.9	Table K.7
		100	0.1	Table K.8
			0.5	Table K.9
			0.9	Table K.10
	2500	10	0.1	Table K.11
			0.5	Table K.12
			0.9	Table K.13
		50	0.1	Table K.14
			0.5	Table K.15
			0.9	Table K.16
		100	0.1	Table K.17
			0.5	Table K.18
			0.9	Table K.19
	5000	10	0.1	Table K.20
			0.5	Table K.21
			0.9	Table K.22
		50	0.1	Table K.23
			0.5	Table K.24
			0.9	Table K.25
		100	0.1	Table K.26
			0.5	Table K.27
			0.9	Table K.28
10000	1000	10	0.1	Table K.29
			0.5	Table K.30
			0.9	Table K.31
		50	0.1	Table K.32
			0.5	Table K.33
			0.9	Table K.34
		100	0.1	Table K.35
			0.5	Table K.36
			0.9	Table K.37
	2500	10	0.1	Table K.38
			0.5	Table K.39
			0.9	Table K.40
		50	0.1	Table K.41
			0.5	Table K.42
			0.9	Table K.43
		100	0.1	Table K.44
			0.5	Table K.45
			0.9	Table K.46
	5000	10	0.1	Table K.47
			0.5	Table K.48
			0.9	Table K.49
		50	0.1	Table K.50
			0.5	Table K.51
			0.9	Table K.52
		100	0.1	Table K.53
			0.5	Table K.54
			0.9	Table K.55

Table K.1: Summary of the simulation results. Please refer to the tables below for the detailed results.

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.81	0.93	0.74	0.98	0.08	0.34	0.05	0.95
0.05	0.50	0.69	1.00	0.55	0.98	0.53	0.99	0.37	0.97
0.05	0.90	0.72	1.00	0.58	0.98	0.71	1.00	0.57	0.98
0.10	0.10	0.86	0.90	0.83	0.97	0.26	0.92	0.16	0.91
0.10	0.50	0.84	1.00	0.74	0.97	0.74	1.00	0.60	0.96
0.10	0.90	0.86	1.00	0.75	0.97	0.85	1.00	0.75	0.97
0.50	0.10	0.94	0.94	0.93	0.94	0.94	0.93	0.94	0.94
0.50	0.50	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
0.50	0.90	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96

Table K.2: Simulation Results: 1000 documents, 1000 words in the vocabulary, 10 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.11	0.39	0.07	0.95	0.02	0.09	0.01	0.95
0.05	0.50	0.23	0.74	0.14	0.96	0.15	0.50	0.09	0.95
0.05	0.90	0.27	0.76	0.17	0.96	0.27	0.74	0.18	0.96
0.10	0.10	0.25	0.58	0.17	0.90	0.10	0.42	0.06	0.90
0.10	0.50	0.39	0.88	0.26	0.92	0.30	0.79	0.19	0.91
0.10	0.90	0.44	0.87	0.30	0.92	0.43	0.84	0.29	0.92
0.50	0.10	0.68	0.69	0.68	0.68	0.69	0.70	0.68	0.69
0.50	0.50	0.77	0.77	0.77	0.77	0.77	0.78	0.77	0.77
0.50	0.90	0.78	0.79	0.78	0.78	0.79	0.79	0.79	0.79

Table K.3: Simulation Results: 1000 documents, 1000 words in the vocabulary, 10 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.06	0.17	0.04	0.95	0.04	0.11	0.03	0.94
0.05	0.50	0.15	0.55	0.09	0.95	0.13	0.33	0.09	0.95
0.05	0.90	0.20	0.53	0.13	0.95	0.20	0.47	0.14	0.95
0.10	0.10	0.17	0.37	0.12	0.89	0.12	0.29	0.08	0.89
0.10	0.50	0.31	0.69	0.21	0.91	0.25	0.53	0.17	0.90
0.10	0.90	0.34	0.65	0.24	0.91	0.33	0.59	0.24	0.91
0.50	0.10	0.62	0.62	0.62	0.62	0.61	0.62	0.62	0.62
0.50	0.50	0.69	0.69	0.70	0.69	0.69	0.70	0.69	0.69
0.50	0.90	0.72	0.72	0.72	0.72	0.71	0.72	0.71	0.71

Table K.4: Simulation Results: 1000 documents, 1000 words in the vocabulary, 10 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	1.00	1.00	1.00	1.00	0.63	0.93	0.52	0.98
0.05	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.05	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.10	1.00	1.00	1.00	1.00	0.91	1.00	0.85	0.98
0.10	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table K.5: Simulation Results: 1000 documents, 1000 words in the vocabulary, 50 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.88	0.94	0.85	0.99	0.11	0.40	0.07	0.95
0.05	0.50	0.89	1.00	0.82	0.99	0.80	1.00	0.69	0.98
0.05	0.90	0.92	1.00	0.85	0.99	0.91	1.00	0.84	0.99
0.10	0.10	0.92	0.93	0.91	0.98	0.37	0.94	0.25	0.93
0.10	0.50	0.95	1.00	0.92	0.99	0.92	1.00	0.86	0.99
0.10	0.90	0.96	1.00	0.93	0.99	0.96	1.00	0.93	0.99
0.50	0.10	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
0.50	0.50	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
0.50	0.90	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99

Table K.6: Simulation Results: 1000 documents, 1000 words in the vocabulary, 50 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.66	0.81	0.59	0.97	0.16	0.57	0.10	0.96
0.05	0.50	0.76	0.99	0.63	0.98	0.62	0.97	0.47	0.97
0.05	0.90	0.79	0.98	0.67	0.98	0.78	0.99	0.66	0.98
0.10	0.10	0.75	0.82	0.71	0.96	0.36	0.91	0.24	0.92
0.10	0.50	0.89	0.99	0.81	0.98	0.82	0.98	0.71	0.97
0.10	0.90	0.90	0.98	0.84	0.98	0.90	0.98	0.83	0.98
0.50	0.10	0.90	0.90	0.91	0.91	0.89	0.89	0.90	0.89
0.50	0.50	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
0.50	0.90	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97

Table K.7: Simulation Results: 1000 documents, 1000 words in the vocabulary, 50 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active F1	Active precision	Active recall	Active accuracy	Passive (Random) F1	Passive (Random) precision	Passive (Random) recall	Passive (Random) accuracy
0.05	0.10	1.00	1.00	1.00	1.00	0.93	0.99	0.90	0.99
0.05	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.05	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table K.8: Simulation Results: 1000 documents, 1000 words in the vocabulary, 100 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active F1	Active precision	Active recall	Active accuracy	Passive (Random) F1	Passive (Random) precision	Passive (Random) recall	Passive (Random) accuracy
0.05	0.10	0.99	1.00	0.99	1.00	0.37	0.76	0.26	0.96
0.05	0.50	0.99	1.00	0.99	1.00	0.98	1.00	0.96	1.00
0.05	0.90	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00
0.10	0.10	0.99	1.00	0.99	1.00	0.75	1.00	0.63	0.96
0.10	0.50	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00
0.10	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table K.9: Simulation Results: 1000 documents, 1000 words in the vocabulary, 100 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.94	0.98	0.92	0.99	0.29	0.75	0.19	0.96
0.05	0.50	0.96	1.00	0.93	1.00	0.92	1.00	0.87	0.99
0.05	0.90	0.97	1.00	0.95	1.00	0.97	1.00	0.95	1.00
0.10	0.10	0.97	0.98	0.97	0.99	0.61	1.00	0.46	0.95
0.10	0.50	0.99	1.00	0.99	1.00	0.98	1.00	0.96	1.00
0.10	0.90	0.99	1.00	0.99	1.00	0.99	1.00	0.99	1.00
0.50	0.10	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99
0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table K.10: Simulation Results: 1000 documents, 1000 words in the vocabulary, 100 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.07	0.26	0.04	0.95	0.00	0.01	0.00	0.95
0.05	0.50	0.07	0.32	0.04	0.95	0.02	0.08	0.01	0.95
0.05	0.90	0.07	0.33	0.04	0.95	0.06	0.29	0.03	0.95
0.10	0.10	0.39	0.76	0.29	0.92	0.02	0.13	0.01	0.90
0.10	0.50	0.32	1.00	0.20	0.92	0.10	0.69	0.06	0.91
0.10	0.90	0.28	1.00	0.17	0.92	0.26	0.98	0.16	0.92
0.50	0.10	0.76	0.76	0.77	0.76	0.74	0.77	0.73	0.75
0.50	0.50	0.83	0.83	0.84	0.84	0.84	0.84	0.84	0.84
0.50	0.90	0.85	0.84	0.85	0.85	0.84	0.84	0.85	0.85

Table K.11: Simulation Results: 1000 documents, 2500 words in the vocabulary, 10 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.00	0.01	0.00	0.95	0.00	0.02	0.00	0.95
0.05	0.50	0.02	0.08	0.01	0.95	0.02	0.08	0.01	0.95
0.05	0.90	0.03	0.14	0.01	0.95	0.02	0.14	0.01	0.95
0.10	0.10	0.03	0.13	0.02	0.89	0.03	0.14	0.02	0.90
0.10	0.50	0.07	0.42	0.04	0.90	0.06	0.31	0.03	0.90
0.10	0.90	0.10	0.45	0.06	0.90	0.09	0.41	0.06	0.90
0.50	0.10	0.57	0.57	0.59	0.57	0.56	0.58	0.56	0.58
0.50	0.50	0.62	0.62	0.63	0.62	0.62	0.62	0.63	0.62
0.50	0.90	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64

Table K.12: Simulation Results: 1000 documents, 2500 words in the vocabulary, 10 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.01	0.04	0.01	0.95	0.01	0.03	0.01	0.95
0.05	0.50	0.02	0.08	0.01	0.95	0.03	0.10	0.02	0.94
0.05	0.90	0.03	0.13	0.02	0.95	0.05	0.13	0.03	0.94
0.10	0.10	0.06	0.17	0.04	0.89	0.05	0.12	0.03	0.89
0.10	0.50	0.08	0.32	0.05	0.90	0.08	0.21	0.05	0.89
0.10	0.90	0.11	0.32	0.07	0.89	0.12	0.30	0.08	0.89
0.50	0.10	0.53	0.54	0.54	0.54	0.53	0.55	0.53	0.54
0.50	0.50	0.57	0.58	0.57	0.58	0.58	0.58	0.58	0.58
0.50	0.90	0.59	0.60	0.59	0.59	0.59	0.60	0.59	0.59

Table K.13: Simulation Results: 1000 documents, 2500 words in the vocabulary, 10 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.94	0.99	0.90	0.99	0.00	0.01	0.00	0.95
0.05	0.50	0.73	1.00	0.59	0.98	0.31	0.83	0.20	0.96
0.05	0.90	0.76	1.00	0.63	0.98	0.72	1.00	0.58	0.98
0.10	0.10	0.95	0.94	0.96	0.99	0.02	0.16	0.01	0.90
0.10	0.50	0.93	1.00	0.87	0.99	0.74	1.00	0.60	0.96
0.10	0.90	0.93	1.00	0.87	0.99	0.92	1.00	0.86	0.98
0.50	0.10	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table K.14: Simulation Results: 1000 documents, 2500 words in the vocabulary, 50 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.06	0.23	0.03	0.95	0.01	0.03	0.00	0.95
0.05	0.50	0.19	0.66	0.11	0.96	0.06	0.30	0.04	0.95
0.05	0.90	0.23	0.73	0.14	0.96	0.21	0.67	0.13	0.96
0.10	0.10	0.35	0.74	0.25	0.92	0.02	0.16	0.01	0.90
0.10	0.50	0.52	0.97	0.37	0.94	0.24	0.94	0.14	0.91
0.10	0.90	0.54	0.99	0.38	0.94	0.50	0.99	0.35	0.93
0.50	0.10	0.77	0.78	0.77	0.77	0.76	0.77	0.77	0.76
0.50	0.50	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89
0.50	0.90	0.91	0.91	0.91	0.91	0.91	0.91	0.92	0.91

Table K.15: Simulation Results: 1000 documents, 2500 words in the vocabulary, 50 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.05	0.22	0.03	0.95	0.02	0.08	0.01	0.95
0.05	0.50	0.12	0.53	0.07	0.95	0.07	0.30	0.04	0.95
0.05	0.90	0.18	0.66	0.10	0.95	0.17	0.67	0.10	0.95
0.10	0.10	0.17	0.53	0.11	0.90	0.06	0.35	0.03	0.90
0.10	0.50	0.40	0.89	0.26	0.92	0.25	0.85	0.15	0.91
0.10	0.90	0.45	0.93	0.30	0.93	0.43	0.90	0.29	0.92
0.50	0.10	0.68	0.68	0.69	0.68	0.68	0.69	0.69	0.68
0.50	0.50	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81
0.50	0.90	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85

Table K.16: Simulation Results: 1000 documents, 2500 words in the vocabulary, 50 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.98	0.98	0.97	1.00	0.01	0.03	0.00	0.95
0.05	0.50	0.98	1.00	0.96	1.00	0.84	1.00	0.74	0.99
0.05	0.90	0.99	1.00	0.98	1.00	0.98	1.00	0.97	1.00
0.10	0.10	1.00	1.00	1.00	1.00	0.12	0.56	0.07	0.91
0.10	0.50	0.99	1.00	0.99	1.00	0.99	1.00	0.98	1.00
0.10	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table K.17: Simulation Results: 1000 documents, 2500 words in the vocabulary, 100 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active F1	Active precision	Active recall	Active accuracy	Passive (Random) F1	Passive (Random) precision	Passive (Random) recall	Passive (Random) accuracy
0.05	0.10	0.55	0.91	0.44	0.97	0.00	0.00	0.00	0.95
0.05	0.50	0.51	0.99	0.37	0.97	0.15	0.58	0.09	0.95
0.05	0.90	0.55	0.99	0.40	0.97	0.51	0.99	0.36	0.97
0.10	0.10	0.76	0.87	0.70	0.96	0.01	0.09	0.00	0.90
0.10	0.50	0.83	1.00	0.71	0.97	0.57	1.00	0.41	0.94
0.10	0.90	0.84	1.00	0.73	0.97	0.82	1.00	0.71	0.97
0.50	0.10	0.91	0.91	0.91	0.91	0.89	0.90	0.89	0.89
0.50	0.50	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
0.50	0.90	0.99	0.98	0.99	0.99	0.98	0.98	0.99	0.99

Table K.18: Simulation Results: 1000 documents, 2500 words in the vocabulary, 100 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active F1	Active precision	Active recall	Active accuracy	Passive (Random) F1	Passive (Random) precision	Passive (Random) recall	Passive (Random) accuracy
0.05	0.10	0.18	0.54	0.11	0.95	0.00	0.02	0.00	0.95
0.05	0.50	0.38	0.92	0.25	0.96	0.16	0.63	0.10	0.95
0.05	0.90	0.42	0.96	0.28	0.96	0.41	0.96	0.27	0.96
0.10	0.10	0.52	0.74	0.42	0.93	0.03	0.24	0.02	0.90
0.10	0.50	0.71	0.99	0.57	0.96	0.43	1.00	0.28	0.93
0.10	0.90	0.72	0.99	0.58	0.96	0.70	0.99	0.55	0.95
0.50	0.10	0.82	0.83	0.82	0.82	0.81	0.82	0.81	0.81
0.50	0.50	0.94	0.94	0.94	0.94	0.93	0.93	0.93	0.93
0.50	0.90	0.95	0.96	0.95	0.96	0.95	0.95	0.95	0.95

Table K.19: Simulation Results: 1000 documents, 2500 words in the vocabulary, 100 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active F1	Active precision	Active recall	Active accuracy	Passive (Random) F1	Passive (Random) precision	Passive (Random) recall	Passive (Random) accuracy
0.05	0.10	0.00	0.01	0.00	0.95	0.00	0.00	0.00	0.95
0.05	0.50	0.00	0.00	0.00	0.95	0.00	0.00	0.00	0.95
0.05	0.90	0.00	0.00	0.00	0.95	0.00	0.00	0.00	0.95
0.10	0.10	0.06	0.36	0.03	0.90	0.01	0.06	0.00	0.90
0.10	0.50	0.05	0.38	0.03	0.90	0.02	0.18	0.01	0.90
0.10	0.90	0.04	0.33	0.02	0.90	0.03	0.30	0.02	0.90
0.50	0.10	0.63	0.64	0.64	0.63	0.64	0.64	0.67	0.64
0.50	0.50	0.71	0.71	0.71	0.71	0.71	0.71	0.72	0.71
0.50	0.90	0.73	0.73	0.74	0.73	0.73	0.73	0.73	0.73

Table K.20: Simulation Results: 1000 documents, 5000 words in the vocabulary, 10 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active F1	Active precision	Active recall	Active accuracy	Passive (Random) F1	Passive (Random) precision	Passive (Random) recall	Passive (Random) accuracy
0.05	0.10	0.00	0.00	0.00	0.95	0.00	0.01	0.00	0.95
0.05	0.50	0.00	0.01	0.00	0.95	0.00	0.02	0.00	0.95
0.05	0.90	0.00	0.00	0.00	0.95	0.01	0.03	0.00	0.95
0.10	0.10	0.02	0.09	0.01	0.90	0.01	0.08	0.01	0.90
0.10	0.50	0.02	0.17	0.01	0.90	0.02	0.10	0.01	0.89
0.10	0.90	0.03	0.20	0.02	0.90	0.03	0.20	0.02	0.89
0.50	0.10	0.52	0.54	0.52	0.54	0.52	0.54	0.53	0.53
0.50	0.50	0.55	0.56	0.55	0.56	0.56	0.56	0.56	0.56
0.50	0.90	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57

Table K.21: Simulation Results: 1000 documents, 5000 words in the vocabulary, 10 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.01	0.02	0.00	0.95	0.01	0.02	0.00	0.95
0.05	0.50	0.01	0.02	0.00	0.94	0.02	0.07	0.01	0.94
0.05	0.90	0.02	0.05	0.01	0.94	0.02	0.08	0.01	0.94
0.10	0.10	0.01	0.05	0.01	0.89	0.02	0.07	0.01	0.89
0.10	0.50	0.03	0.15	0.02	0.89	0.05	0.13	0.03	0.88
0.10	0.90	0.05	0.16	0.03	0.88	0.06	0.16	0.04	0.88
0.50	0.10	0.52	0.52	0.53	0.52	0.51	0.52	0.52	0.52
0.50	0.50	0.53	0.54	0.54	0.54	0.53	0.53	0.53	0.53
0.50	0.90	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54

Table K.22: Simulation Results: 1000 documents, 5000 words in the vocabulary, 10 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.26	0.65	0.18	0.96	0.00	0.00	0.00	0.95
0.05	0.50	0.02	0.13	0.01	0.95	0.00	0.00	0.00	0.95
0.05	0.90	0.03	0.14	0.01	0.95	0.02	0.12	0.01	0.95
0.10	0.10	0.68	0.82	0.67	0.94	0.00	0.00	0.00	0.90
0.10	0.50	0.38	1.00	0.24	0.92	0.03	0.25	0.02	0.90
0.10	0.90	0.31	0.99	0.19	0.92	0.25	0.96	0.15	0.92
0.50	0.10	0.86	0.88	0.87	0.86	0.86	0.90	0.85	0.86
0.50	0.50	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
0.50	0.90	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98

Table K.23: Simulation Results: 1000 documents, 5000 words in the vocabulary, 50 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.00	0.01	0.00	0.95	0.00	0.00	0.00	0.95
0.05	0.50	0.00	0.02	0.00	0.95	0.00	0.00	0.00	0.95
0.05	0.90	0.01	0.04	0.00	0.95	0.01	0.04	0.00	0.95
0.10	0.10	0.02	0.12	0.01	0.90	0.00	0.05	0.00	0.90
0.10	0.50	0.04	0.41	0.02	0.90	0.02	0.17	0.01	0.90
0.10	0.90	0.06	0.52	0.03	0.90	0.06	0.45	0.03	0.90
0.50	0.10	0.62	0.64	0.63	0.63	0.60	0.64	0.60	0.63
0.50	0.50	0.73	0.73	0.74	0.73	0.73	0.74	0.73	0.74
0.50	0.90	0.77	0.77	0.78	0.77	0.77	0.77	0.77	0.77

Table K.24: Simulation Results: 1000 documents, 5000 words in the vocabulary, 50 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.01	0.04	0.00	0.95	0.02	0.08	0.01	0.95
0.05	0.50	0.01	0.05	0.01	0.95	0.02	0.07	0.01	0.95
0.05	0.90	0.02	0.09	0.01	0.95	0.02	0.09	0.01	0.95
0.10	0.10	0.03	0.18	0.02	0.90	0.04	0.16	0.03	0.90
0.10	0.50	0.05	0.36	0.03	0.90	0.06	0.40	0.03	0.90
0.10	0.90	0.09	0.47	0.05	0.90	0.09	0.46	0.05	0.90
0.50	0.10	0.58	0.58	0.59	0.59	0.57	0.58	0.58	0.58
0.50	0.50	0.66	0.66	0.66	0.66	0.66	0.66	0.67	0.66
0.50	0.90	0.70	0.69	0.70	0.70	0.69	0.70	0.70	0.70

Table K.25: Simulation Results: 1000 documents, 5000 words in the vocabulary, 50 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.86	0.97	0.81	0.99	0.00	0.00	0.00	0.95
0.05	0.50	0.21	0.73	0.13	0.96	0.01	0.02	0.00	0.95
0.05	0.90	0.24	0.75	0.15	0.96	0.18	0.63	0.11	0.96
0.10	0.10	0.89	0.90	0.90	0.98	0.00	0.00	0.00	0.90
0.10	0.50	0.79	1.00	0.66	0.97	0.23	0.88	0.14	0.92
0.10	0.90	0.77	1.00	0.64	0.97	0.73	1.00	0.59	0.96
0.50	0.10	0.95	0.95	0.95	0.95	0.96	0.97	0.96	0.96
0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table K.26: Simulation Results: 1000 documents, 5000 words in the vocabulary, 100 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.00	0.01	0.00	0.95	0.00	0.00	0.00	0.95
0.05	0.50	0.01	0.08	0.01	0.95	0.00	0.01	0.00	0.95
0.05	0.90	0.03	0.16	0.02	0.95	0.03	0.13	0.01	0.95
0.10	0.10	0.11	0.51	0.07	0.90	0.00	0.00	0.00	0.90
0.10	0.50	0.21	0.91	0.12	0.91	0.03	0.27	0.02	0.90
0.10	0.90	0.22	0.92	0.13	0.91	0.17	0.84	0.10	0.91
0.50	0.10	0.73	0.75	0.73	0.74	0.70	0.74	0.72	0.72
0.50	0.50	0.88	0.88	0.88	0.88	0.88	0.87	0.88	0.88
0.50	0.90	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91

Table K.27: Simulation Results: 1000 documents, 5000 words in the vocabulary, 100 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.01	0.02	0.00	0.95	0.00	0.00	0.00	0.95
0.05	0.50	0.01	0.06	0.01	0.95	0.01	0.04	0.00	0.95
0.05	0.90	0.03	0.14	0.02	0.95	0.03	0.15	0.02	0.95
0.10	0.10	0.05	0.31	0.03	0.90	0.01	0.07	0.00	0.90
0.10	0.50	0.17	0.84	0.09	0.91	0.05	0.36	0.03	0.90
0.10	0.90	0.19	0.88	0.11	0.91	0.17	0.86	0.09	0.91
0.50	0.10	0.65	0.66	0.65	0.65	0.65	0.66	0.66	0.65
0.50	0.50	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78
0.50	0.90	0.82	0.83	0.82	0.82	0.82	0.83	0.82	0.83

Table K.28: Simulation Results: 1000 documents, 5000 words in the vocabulary, 100 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.88	1.00	0.78	0.99	0.60	1.00	0.44	0.97
0.05	0.50	0.91	1.00	0.84	0.99	0.90	1.00	0.82	0.99
0.05	0.90	0.93	1.00	0.87	0.99	0.93	1.00	0.87	0.99
0.10	0.10	0.92	1.00	0.86	0.99	0.78	1.00	0.64	0.96
0.10	0.50	0.93	1.00	0.87	0.99	0.93	1.00	0.87	0.99
0.10	0.90	0.93	1.00	0.87	0.99	0.93	1.00	0.87	0.99
0.50	0.10	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
0.50	0.50	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
0.50	0.90	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93

Table K.29: Simulation Results: 10000 documents, 1000 words in the vocabulary, 10 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.65	0.99	0.49	0.97	0.25	0.78	0.15	0.96
0.05	0.50	0.65	0.98	0.49	0.97	0.59	0.90	0.44	0.97
0.05	0.90	0.67	0.95	0.52	0.97	0.66	0.93	0.52	0.97
0.10	0.10	0.69	0.98	0.53	0.95	0.43	0.81	0.29	0.92
0.10	0.50	0.69	0.98	0.54	0.95	0.67	0.90	0.53	0.95
0.10	0.90	0.69	0.96	0.54	0.95	0.69	0.94	0.54	0.95
0.50	0.10	0.76	0.76	0.76	0.76	0.77	0.77	0.77	0.77
0.50	0.50	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77
0.50	0.90	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77

Table K.30: Simulation Results: 10000 documents, 1000 words in the vocabulary, 10 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.52	0.97	0.36	0.97	0.21	0.50	0.14	0.95
0.05	0.50	0.53	0.97	0.37	0.97	0.47	0.80	0.33	0.96
0.05	0.90	0.56	0.93	0.40	0.97	0.55	0.88	0.40	0.97
0.10	0.10	0.56	0.91	0.41	0.94	0.34	0.59	0.24	0.91
0.10	0.50	0.59	0.97	0.42	0.94	0.56	0.82	0.42	0.93
0.10	0.90	0.59	0.92	0.43	0.94	0.58	0.90	0.43	0.94
0.50	0.10	0.70	0.70	0.70	0.70	0.70	0.71	0.70	0.71
0.50	0.50	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72
0.50	0.90	0.72	0.72	0.71	0.72	0.72	0.72	0.72	0.72

Table K.31: Simulation Results: 10000 documents, 1000 words in the vocabulary, 10 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active F1	Active precision	Active recall	Active accuracy	Passive (Random) F1	Passive (Random) precision	Passive (Random) recall	Passive (Random) accuracy
0.05	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.05	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.05	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table K.32: Simulation Results: 10000 documents, 1000 words in the vocabulary, 50 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active F1	Active precision	Active recall	Active accuracy	Passive (Random) F1	Passive (Random) precision	Passive (Random) recall	Passive (Random) accuracy
0.05	0.10	0.97	1.00	0.94	1.00	0.90	1.00	0.83	0.99
0.05	0.50	0.98	1.00	0.97	1.00	0.98	1.00	0.97	1.00
0.05	0.90	0.99	1.00	0.98	1.00	0.99	1.00	0.98	1.00
0.10	0.10	0.98	1.00	0.97	1.00	0.96	1.00	0.92	0.99
0.10	0.50	0.99	1.00	0.98	1.00	0.99	1.00	0.98	1.00
0.10	0.90	0.99	1.00	0.98	1.00	0.99	1.00	0.98	1.00
0.50	0.10	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
0.50	0.50	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
0.50	0.90	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99

Table K.33: Simulation Results: 10000 documents, 1000 words in the vocabulary, 50 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.94	0.99	0.89	0.99	0.82	0.99	0.70	0.98
0.05	0.50	0.96	1.00	0.92	1.00	0.95	0.99	0.91	1.00
0.05	0.90	0.96	0.99	0.94	1.00	0.96	0.99	0.94	1.00
0.10	0.10	0.95	0.98	0.93	0.99	0.90	0.98	0.83	0.98
0.10	0.50	0.97	0.99	0.95	0.99	0.97	0.99	0.95	0.99
0.10	0.90	0.97	0.99	0.95	0.99	0.97	0.99	0.95	0.99
0.50	0.10	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
0.50	0.50	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
0.50	0.90	0.97	0.97	0.98	0.97	0.97	0.97	0.97	0.97

Table K.34: Simulation Results: 10000 documents, 1000 words in the vocabulary, 50 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.05	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.05	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table K.35: Simulation Results: 10000 documents, 1000 words in the vocabulary, 100 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
0.05	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.05	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table K.36: Simulation Results: 10000 documents, 1000 words in the vocabulary, 100 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.99	1.00	0.98	1.00	0.98	1.00	0.95	1.00
0.05	0.50	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00
0.05	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.10	1.00	1.00	0.99	1.00	0.99	1.00	0.99	1.00
0.10	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table K.37: Simulation Results: 10000 documents, 1000 words in the vocabulary, 100 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.88	1.00	0.78	0.99	0.60	1.00	0.44	0.97
0.05	0.50	0.91	1.00	0.84	0.99	0.90	1.00	0.82	0.99
0.05	0.90	0.93	1.00	0.87	0.99	0.93	1.00	0.87	0.99
0.10	0.10	0.92	1.00	0.86	0.99	0.78	1.00	0.64	0.96
0.10	0.50	0.93	1.00	0.87	0.99	0.93	1.00	0.87	0.99
0.10	0.90	0.93	1.00	0.87	0.99	0.93	1.00	0.87	0.99
0.50	0.10	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
0.50	0.50	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
0.50	0.90	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93

Table K.38: Simulation Results: 10000 documents, 2500 words in the vocabulary, 10 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.65	0.99	0.49	0.97	0.25	0.78	0.15	0.96
0.05	0.50	0.65	0.98	0.49	0.97	0.59	0.90	0.44	0.97
0.05	0.90	0.67	0.95	0.52	0.97	0.66	0.93	0.52	0.97
0.10	0.10	0.69	0.98	0.53	0.95	0.43	0.81	0.29	0.92
0.10	0.50	0.69	0.98	0.54	0.95	0.67	0.90	0.53	0.95
0.10	0.90	0.69	0.96	0.54	0.95	0.69	0.94	0.54	0.95
0.50	0.10	0.76	0.76	0.76	0.76	0.77	0.77	0.77	0.77
0.50	0.50	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77
0.50	0.90	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77

Table K.39: Simulation Results: 10000 documents, 2500 words in the vocabulary, 10 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.52	0.97	0.36	0.97	0.21	0.50	0.14	0.95
0.05	0.50	0.53	0.97	0.37	0.97	0.47	0.80	0.33	0.96
0.05	0.90	0.56	0.93	0.40	0.97	0.55	0.88	0.40	0.97
0.10	0.10	0.56	0.91	0.41	0.94	0.34	0.59	0.24	0.91
0.10	0.50	0.59	0.97	0.42	0.94	0.56	0.82	0.42	0.93
0.10	0.90	0.59	0.92	0.43	0.94	0.58	0.90	0.43	0.94
0.50	0.10	0.70	0.70	0.70	0.70	0.70	0.71	0.70	0.71
0.50	0.50	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72
0.50	0.90	0.72	0.72	0.71	0.72	0.72	0.72	0.72	0.72

Table K.40: Simulation Results: 10000 documents, 2500 words in the vocabulary, 10 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.05	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.05	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table K.41: Simulation Results: 10000 documents, 2500 words in the vocabulary, 50 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active F1	Active precision	Active recall	Active accuracy	Passive (Random) F1	Passive (Random) precision	Passive (Random) recall	Passive (Random) accuracy
0.05	0.10	0.97	1.00	0.94	1.00	0.90	1.00	0.83	0.99
0.05	0.50	0.98	1.00	0.97	1.00	0.98	1.00	0.97	1.00
0.05	0.90	0.99	1.00	0.98	1.00	0.99	1.00	0.98	1.00
0.10	0.10	0.98	1.00	0.97	1.00	0.96	1.00	0.92	0.99
0.10	0.50	0.99	1.00	0.98	1.00	0.99	1.00	0.98	1.00
0.10	0.90	0.99	1.00	0.98	1.00	0.99	1.00	0.98	1.00
0.50	0.10	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
0.50	0.50	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
0.50	0.90	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99

Table K.42: Simulation Results: 10000 documents, 2500 words in the vocabulary, 50 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active F1	Active precision	Active recall	Active accuracy	Passive (Random) F1	Passive (Random) precision	Passive (Random) recall	Passive (Random) accuracy
0.05	0.10	0.94	0.99	0.89	0.99	0.82	0.99	0.70	0.98
0.05	0.50	0.96	1.00	0.92	1.00	0.95	0.99	0.91	1.00
0.05	0.90	0.96	0.99	0.94	1.00	0.96	0.99	0.94	1.00
0.10	0.10	0.95	0.98	0.93	0.99	0.90	0.98	0.83	0.98
0.10	0.50	0.97	0.99	0.95	0.99	0.97	0.99	0.95	0.99
0.10	0.90	0.97	0.99	0.95	0.99	0.97	0.99	0.95	0.99
0.50	0.10	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
0.50	0.50	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
0.50	0.90	0.97	0.97	0.98	0.97	0.97	0.97	0.97	0.97

Table K.43: Simulation Results: 10000 documents, 2500 words in the vocabulary, 50 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00
0.05	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.05	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table K.44: Simulation Results: 10000 documents, 2500 words in the vocabulary, 100 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.96	1.00	0.92	1.00	0.65	1.00	0.49	0.97
0.05	0.50	0.97	1.00	0.94	1.00	0.97	1.00	0.94	1.00
0.05	0.90	0.98	1.00	0.96	1.00	0.98	1.00	0.97	1.00
0.10	0.10	0.97	0.99	0.95	0.99	0.88	1.00	0.78	0.98
0.10	0.50	0.99	1.00	0.98	1.00	0.98	1.00	0.97	1.00
0.10	0.90	0.99	0.99	0.98	1.00	0.99	1.00	0.98	1.00
0.50	0.10	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
0.50	0.50	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
0.50	0.90	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99

Table K.45: Simulation Results: 10000 documents, 2500 words in the vocabulary, 100 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.89	0.97	0.83	0.99	0.46	1.00	0.30	0.97
0.05	0.50	0.92	0.99	0.86	0.99	0.90	0.98	0.82	0.99
0.05	0.90	0.94	0.99	0.89	0.99	0.93	0.98	0.89	0.99
0.10	0.10	0.92	0.95	0.89	0.98	0.73	0.99	0.58	0.96
0.10	0.50	0.95	0.99	0.92	0.99	0.94	0.98	0.90	0.99
0.10	0.90	0.96	0.98	0.93	0.99	0.96	0.98	0.93	0.99
0.50	0.10	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
0.50	0.50	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
0.50	0.90	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97

Table K.46: Simulation Results: 10000 documents, 2500 words in the vocabulary, 100 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.33	1.00	0.20	0.96	0.00	0.20	0.00	0.95
0.05	0.50	0.24	1.00	0.14	0.96	0.14	1.00	0.07	0.95
0.05	0.90	0.29	0.99	0.17	0.96	0.28	0.99	0.16	0.96
0.10	0.10	0.57	0.96	0.41	0.94	0.03	0.93	0.01	0.90
0.10	0.50	0.47	1.00	0.31	0.93	0.32	0.99	0.19	0.92
0.10	0.90	0.52	0.99	0.36	0.94	0.50	0.99	0.34	0.93
0.50	0.10	0.72	0.72	0.73	0.72	0.73	0.73	0.73	0.73
0.50	0.50	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74
0.50	0.90	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74

Table K.47: Simulation Results: 10000 documents, 5000 words in the vocabulary, 10 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.02	0.64	0.01	0.95	0.01	0.10	0.00	0.95
0.05	0.50	0.06	0.73	0.03	0.95	0.06	0.37	0.03	0.95
0.05	0.90	0.11	0.56	0.06	0.95	0.11	0.49	0.06	0.95
0.10	0.10	0.13	0.70	0.07	0.90	0.03	0.24	0.02	0.90
0.10	0.50	0.16	0.80	0.09	0.91	0.14	0.46	0.08	0.90
0.10	0.90	0.22	0.63	0.13	0.90	0.21	0.56	0.13	0.90
0.50	0.10	0.57	0.57	0.57	0.57	0.57	0.57	0.58	0.57
0.50	0.50	0.59	0.59	0.59	0.59	0.59	0.60	0.59	0.60
0.50	0.90	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60

Table K.48: Simulation Results: 10000 documents, 5000 words in the vocabulary, 10 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.02	0.32	0.01	0.95	0.02	0.08	0.01	0.94
0.05	0.50	0.05	0.56	0.03	0.95	0.06	0.20	0.04	0.94
0.05	0.90	0.09	0.34	0.05	0.95	0.09	0.28	0.05	0.95
0.10	0.10	0.09	0.42	0.05	0.90	0.07	0.16	0.04	0.88
0.10	0.50	0.12	0.56	0.07	0.90	0.12	0.27	0.08	0.89
0.10	0.90	0.16	0.41	0.10	0.90	0.16	0.34	0.11	0.89
0.50	0.10	0.54	0.54	0.54	0.54	0.54	0.54	0.55	0.54
0.50	0.50	0.56	0.56	0.56	0.56	0.56	0.56	0.57	0.56
0.50	0.90	0.57	0.56	0.57	0.57	0.57	0.57	0.57	0.57

Table K.49: Simulation Results: 10000 documents, 5000 words in the vocabulary, 10 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.90	1.00	0.81	0.99	0.03	0.71	0.02	0.95
0.05	0.50	0.90	1.00	0.81	0.99	0.79	1.00	0.66	0.98
0.05	0.90	0.91	1.00	0.83	0.99	0.91	1.00	0.83	0.99
0.10	0.10	0.95	1.00	0.90	0.99	0.29	1.00	0.17	0.92
0.10	0.50	0.94	1.00	0.90	0.99	0.92	1.00	0.85	0.99
0.10	0.90	0.96	1.00	0.93	0.99	0.97	1.00	0.93	0.99
0.50	0.10	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
0.50	0.50	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
0.50	0.90	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98

Table K.50: Simulation Results: 10000 documents, 5000 words in the vocabulary, 50 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.43	0.89	0.28	0.96	0.01	0.43	0.01	0.95
0.05	0.50	0.52	0.97	0.36	0.97	0.36	0.95	0.22	0.96
0.05	0.90	0.56	0.95	0.40	0.97	0.54	0.93	0.38	0.97
0.10	0.10	0.56	0.75	0.45	0.93	0.08	0.91	0.04	0.90
0.10	0.50	0.70	0.94	0.56	0.95	0.58	0.91	0.43	0.94
0.10	0.90	0.72	0.92	0.59	0.95	0.70	0.90	0.58	0.95
0.50	0.10	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78
0.50	0.50	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83
0.50	0.90	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83

Table K.51: Simulation Results: 10000 documents, 5000 words in the vocabulary, 50 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.17	0.79	0.10	0.95	0.03	0.42	0.01	0.95
0.05	0.50	0.36	0.90	0.23	0.96	0.26	0.76	0.16	0.96
0.05	0.90	0.42	0.81	0.29	0.96	0.41	0.78	0.28	0.96
0.10	0.10	0.34	0.63	0.24	0.91	0.10	0.60	0.06	0.90
0.10	0.50	0.53	0.85	0.39	0.93	0.44	0.75	0.31	0.92
0.10	0.90	0.57	0.81	0.44	0.93	0.55	0.77	0.43	0.93
0.50	0.10	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69
0.50	0.50	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76
0.50	0.90	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76

Table K.52: Simulation Results: 10000 documents, 5000 words in the vocabulary, 50 words per document, β=0.9

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.99	1.00	0.97	1.00	0.24	1.00	0.14	0.96
0.05	0.50	0.99	1.00	0.97	1.00	0.98	1.00	0.97	1.00
0.05	0.90	0.99	1.00	0.98	1.00	0.99	1.00	0.99	1.00
0.10	0.10	0.99	1.00	0.99	1.00	0.77	1.00	0.64	0.96
0.10	0.50	0.99	1.00	0.99	1.00	1.00	1.00	0.99	1.00
0.10	0.90	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
0.50	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.50	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table K.53: Simulation Results: 10000 documents, 5000 words in the vocabulary, 100 words per document, β=0.1

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.78	0.94	0.67	0.98	0.03	0.75	0.02	0.95
0.05	0.50	0.81	0.99	0.69	0.98	0.69	0.99	0.53	0.98
0.05	0.90	0.83	0.99	0.72	0.99	0.83	0.99	0.71	0.99
0.10	0.10	0.82	0.88	0.77	0.97	0.23	1.00	0.13	0.91
0.10	0.50	0.88	0.98	0.81	0.98	0.84	0.98	0.74	0.97
0.10	0.90	0.90	0.98	0.84	0.98	0.90	0.97	0.84	0.98
0.50	0.10	0.92	0.92	0.91	0.92	0.91	0.91	0.91	0.91
0.50	0.50	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
0.50	0.90	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94

Table K.54: Simulation Results: 10000 documents, 5000 words in the vocabulary, 100 words per document, β=0.5

Prop. Positive	Prop. Hand-label	Active				Passive (Random)			
		F1	precision	recall	accuracy	F1	precision	recall	accuracy
0.05	0.10	0.56	0.81	0.43	0.97	0.04	0.79	0.02	0.95
0.05	0.50	0.67	0.97	0.52	0.98	0.51	0.95	0.35	0.97
0.05	0.90	0.70	0.95	0.55	0.98	0.68	0.94	0.54	0.98
0.10	0.10	0.64	0.75	0.55	0.94	0.18	0.94	0.10	0.91
0.10	0.50	0.79	0.94	0.68	0.96	0.71	0.92	0.58	0.95
0.10	0.90	0.80	0.93	0.71	0.97	0.80	0.92	0.70	0.96
0.50	0.10	0.83	0.83	0.83	0.83	0.82	0.82	0.82	0.82
0.50	0.50	0.89	0.89	0.89	0.89	0.88	0.88	0.89	0.88
0.50	0.90	0.89	0.88	0.89	0.89	0.89	0.89	0.89	0.89

Table K.55: Simulation Results: 10000 documents, 5000 words in the vocabulary, 100 words per document, β=0.9

References

Gohdes, A. (2020), "Repression technology: Internet accessibility and state violence," *American Journal of Political Science*, 64(3), 488–503.

Miller, B., Linder, F., and Mebane, W. R. (2020), "Active Learning Approaches for Labeling Text: Review and Assessment of the Performance of Active Learning Approaches," *Political Analysis*, pp. 1–20.

Park, B., Greene, K., and Colaresi, M. (2020), "Human Rights are (Increasingly) Plural: Learning the Changing Taxonomy of Human Rights from Large-scale Text Reveals Information Effects," *American Political Science Review*, 114(3), 888–910.