Bill Survival: American Legislative Action on AI

Data Science for Public Policy

Emil Fahrig
ETH Zürich

Noah Fehr
ETH Zürich

David Hollenstein
ETH Zürich

July 2025

1 Introduction

Legislation traditionally lags behind technological change, and the rapid rise of artificial intelligence (AI) embodies this pattern. Governments across jurisdictions have scrambled to respond to this increasingly disruptive series of innovations. The public release of tools like ChatGPT has forced legislative action, raising broader questions over the efficacy and appropriate nature of such legislation. This proliferation of legislative activity informs key questions of lawmakers' response to emerging technologies and the subsequent legislative outcomes. This study explores the survival of AI-related bills across branches in the US to understand when and how legislators choose to regulate or support innovation.

Understanding these legislative dynamics is vital both to contextualize current regulatory action and to anticipate how governments will address the societal and ethical challenges of emerging technologies. As AI becomes further integrated into daily life, the timing and nature of regulatory interventions will influence public trust, economic development, and civil liberties. This paper aims to investigate this by employing natural language processing to analyze U.S. legislative acts related to AI and assess lawmakers' interventions across different applications of these technologies.

This project aims to answer the following question:
"To what extent, and under what circumstances, do subtopics of AI legislation affect the survival of such bills in the U.S.?"

The focus on the U.S. is grounded in the innovation leadership of the U.S. in artificial intelligence and the legislative diversity due to a federalist system. Since we focus exclusively on a single country, variation in our dataset arises from both time and the authority responsible for the legislation (e.g., federal or state, legislative or executive). Each observation represents a single AI-related legislative act. Our dependent variable is bill enactment. We hypothesize that internal alignment within an authority increases a bill's likelihood of passing, while disagreement lowers it. Furthermore, examining the content of successful bills may reveal how legislators frame and prioritize certain aspects and applications of AI.

We base our analysis on the AGORA dataset from the Emerging Technology Observatory, which compiles AI-related laws, regulations, and governance documents (Haddii, 2025). Our approach includes three steps:

1. **Text Preprocessing:** We clean the legislative texts, removing standard stop words and extracting legal-specific stop words—terms frequent in legislative writing but low in substantive meaning. These are generated using an expanded word embedding dictionary.

2. **Topic Modeling:** Using BERTopic, we extract the main themes from legislative texts. Each document receives a topic distribution score from 0 to 1, indicating its pertinence to each topic. These topic scores serve as the main independent variables for our subsequent analysis. We identify eight broad topics:

   - Defense & Education
   - AI & Cybersecurity

- Energy & Technology
- Media & Privacy
- Automated Decision Systems
- Foreign Affairs
- Healthcare Services
- Government & Data Agencies

3. **Statistical Modeling:** We build a logit regression model with bill survival as the dependent variable, incorporating the AGORA dataset variables and the topic scores. We begin with a baseline model, then include time- and authority-fixed effects and interaction terms.

Our findings suggest three main conclusions: First, alignment within a legislative authority appears to decrease the likelihood of bill passage, likely due to reduced debate. Second, bills concerning the private sector are less likely to survive than those focused on the public sphere. Finally, the impact of a bill's topical focus is time-sensitive, potentially indicating the rise and fall of divisiveness on certain issues.

2 Literature Review

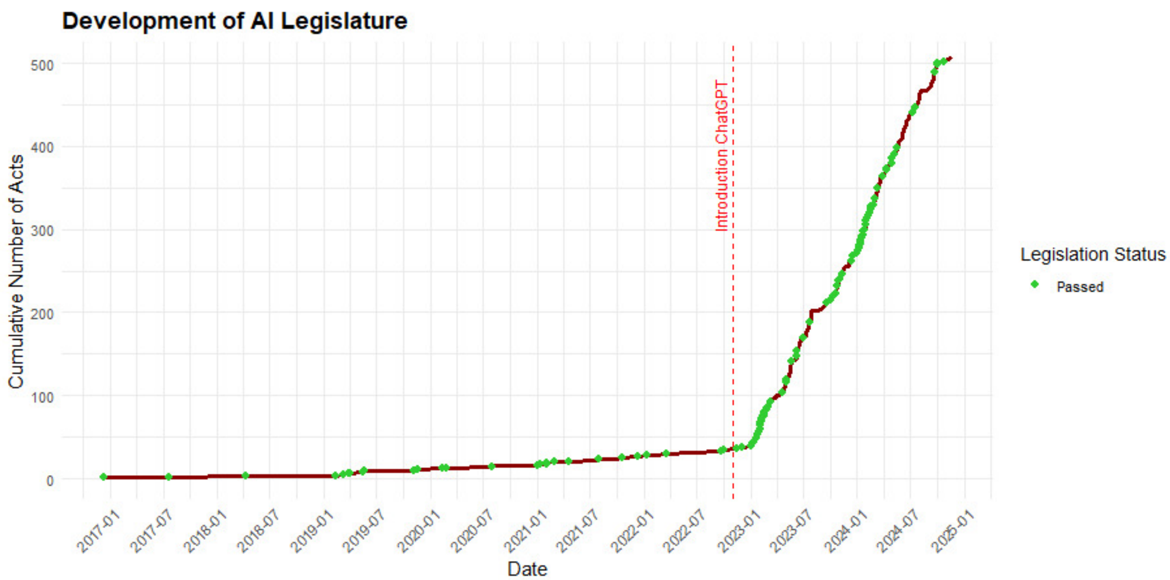
The research question stated above is inherently rooted in the literature on bill survival. Recent scholarship has increasingly examined the role of bill topic in predicting legislative success within the U.S. Congress. Traditional analyses have emphasized sponsor characteristics and institutional dynamics, but contemporary studies highlight the predictive power of bill content. For instance, Nay (2016) utilized machine learning to analyze over 68,000 bills from 2001 to 2015, finding that specific language within bills significantly influences their likelihood of enactment. Terms like "impact" and "effects" were associated with higher success rates in climate-related legislation, whereas words such as "global" and "warming" correlated with lower passage probabilities. Similarly, Adler and Wilkerson (2005) observed that bills addressing urgent or uncontroversial issues tend to have higher success rates, suggesting that the nature of the policy topic itself can affect legislative outcomes. Building on this, Yano (2012) developed models that predicted whether bills would survive the opaque committee stage by incorporating both sponsor-related and textual features, showing that functional categories such as "important" or "recurring" policy issues significantly enhanced prediction accuracy. Likewise, Eidelman et al. (2018) demonstrated that combining contextual information about sponsors, committees, and legislative structure with the lexical content of bills improves predictions of floor action across all 50 U.S. states, with accuracy reaching up to 86%. These findings collectively suggest that the subject matter and textual framing of legislation are powerful predictors of its survival, reinforcing a broader shift toward content-based analysis in understanding legislative success. Our analysis builds upon this literature and focuses on the mechanisms of causal inference. The topical relevance for prediction informed our approach to understanding the factors which have contributed to bill enactment.

3 Data and Summary Statistics

The requirements for the dataset to use for this project were twofold: First, we needed the full texts of each legislative act in order to make our inferences. Executive summaries or intended target were insufficient since they introduced an intermediate entity that might mis- or re-interpret the content of the acts. This was crucial since the analysis targeted topics mentioned and semantics used. Second, the topic variables that are generated through the BERTopic are necessary, but insufficient to answer the broader questions stated above. The dataset therefore needs to contain meta-variables on the legislature such as geographic location, authority as well as our dependent variable: whether the legislation did survive the legislative process and was eventually signed into law.

The AI Governance and Regulatory Archive (AGORA) dataset of the Emerging Technologies Observatory summarises this information over a total of 650 observations. At its core, this dataset is a collection of legislative texts from the US (with selected texts from other entities such as the UN) on the topic of artificial intelligence. It contains 101 meta variables on these texts, of which we modify the following slightly:

The dependent variable "Most recent activity" has three realisations: "Proposed", "Enacted" as well as "Defunct". We summarise "Enacted" and "Defunct" into one category since the latter means that the legislation was in effect at one point but has since become obsolete. This leaves the dataset with 296 enacted and 354 not enacted bills. The following plot illustrates the cumulative number of acts as well as their status in the raw dataset, reiterating the growing interest in and legislative effort regarding artificial intelligence:



We further slightly modify the variable capturing the responsible authority slightly: Individual expressions for each legislative or executive branch on the federal or state level are summarised into the following categories: "Red state", "Blue state", "Purple state" as well as "Federal legislative" and "Federal executive" branch. The latter is used as a reference category for this variable since we anticipate acts by the federal executive branch to face the least resistance in enactment.

4 Methods

4.1 Natural Language Processing

In this project, we calculate a distribution over extracted topics for each document in the AGORA dataset. We use NLP methods for two tasks: to remove uninformative legalese from documents and to obtain a distribution over topics for each document in the AGORA dataset.

4.1.1 Removing Legalese

First, we remove legalese from the legislative text. When performing LDA or BERTopic on the raw text, the generated topics contained legal words (i.e., 'bill' or 'statute') instead of focusing on content-related themes. In order to create more informative topics for subsequent analysis, we work to remove some of these noisy legal terms. We start with a base dictionary of 21 common legal words that often appear in the document and expand this using a word2vec word embedding model. Word2vec uses a classic neural network architecture with three layers:

1. V-dimensional input layer takas a one-hot-encoded representation of a word(s) where V is the size of the vocabulary that word2vec considers.

2. 300-dimensional hidden layer consisting of linear neurons, which means that they use the identity activation function.

3. V-dimensional final layer which uses a SoftMax activation function.

Prior to training the neural net, we preprocess the corpus by setting all characters to lowercase, removing punctuation, splitting text into tokens, removing stop words, replacing digits with #, and stemming words. The preprocessed corpus can now be passed to word2vec. Word2vec is trained with either the skip-gram objective or the continuous bag-of-words (CBOW) objective; by default, genism uses CBOW which attempts to predict the center word given the context words within the window size. To achieve this, all context words are one-hot-encoded and summed. The resulting vector is fed into the neural network and we receive a probability distribution over the vocabulary. Next, the cross-entropy loss is used to penalize low probabilities assigned to the center word. The loss function is given below:

$$L(\theta) = -\sum_{i=0}^{|V|} y_i log(\hat{y}_i) \tag{1}$$

Where y_i is the i^{th} element in the one-hot-encoded center word and $\hat{y}_i$ is the i^{th} element in the outputted probability distribution from the word2vec model. After the neural network is trained to a point where the loss is plateauing, the hidden layer activations are used as the embeddings for each word. Those embeddings are stored in the columns of the matrix which transforms the input into the hidden layer. After training, we obtain a word embedding for each word mentioned in the corpus of documents at least 25 times. Subsequently, the nearest five words of each base word (using cosine similarity) are added to the list of legal words. In the end, we have a list of 126 legal words which are removed from the dataset.

4.1.2 Extracting Content-Related Topics

The second NLP application is the use of a BERTopic model to extract latent topics from the preprocessed corpus. BERTopic is a pipeline of steps, namely:

1. **Embeddings the documents by sentence using Sentence BERT**
   Sentence BERT is a pre-trained model which uses a contrastive loss to penalize a low cosine similarity between the sentence embeddings of similar sentences and penalize a high cosine similarity between the sentence embeddings of dissimilar sentences. Datasets like the Stanford Natural Language Inference (SNLI) or the Multi-Genre Natural Language Inference (MNLI) were used to pre-train Sentence BERT as they provide sentence pairs and their respective similarity scores. The training process is rather cumbersome. Tokenized sentence pairs are separately embedded using BERT; the embeddings for all tokens in both sentences are pooled to create a sentence embedding.

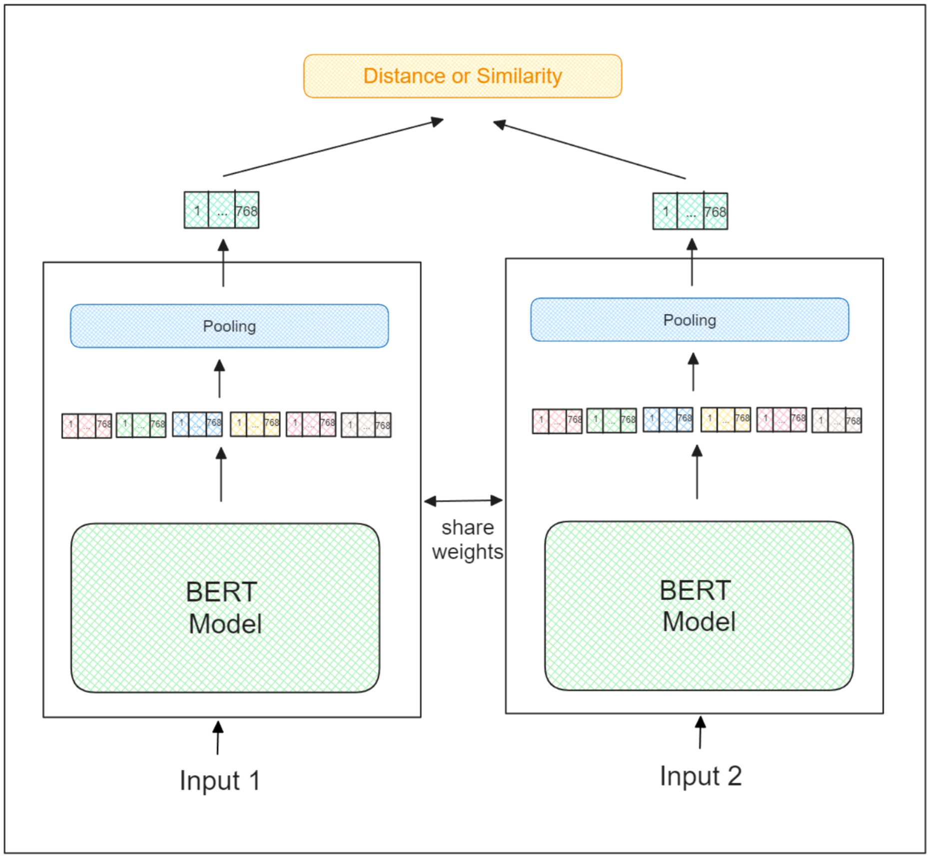


Figure 1: Siamese Sentence BERT architectural diagram for pre-training

A BERT-based tokenizer is used for sentence tokenization while average pooling is employed to combine the individual token embeddings. Once this step is finished for all sentence pairs, the Multiple Negative Ranking Loss (MNRL)

$$MNRL = -\sum_{i=1}^{N} \log \frac{e^{sim(a_i,p_i)}}{\sum_{j=1}^{N} e^{sim(a_i,p_j)}} \tag{2}$$

loss function is used to quantify the quality of the embeddings where $sim(a, b)$ is the cosine distance between vectors a and b. Anchor sentence a_i and positive sentence p_i is a pair which was simultaneously passed through the Sentence BERT model shown above since they have a high similarity score in the dataset. All other positive sentences are deemed to be negative examples of a_i; pairs for which we would like the similarity score between a_i and p_j to be low. In the loss seen above, for each anchor a_i, we create a ratio between $sim(a_i, p_i)$ and $\sum_{j=1}^{N} sim(a_i, p_j)$. This is a probability so values range in the interval $[0, 1]$. Ideally, the value is 1 since that would indicate $sim(a_i, p_i)$ is non-zero and $sim(a_i, p_j)\forall j$ is 0 meaning $\log \frac{e^{sim(a_i,p_i)}}{\sum_{j=1}^{N} e^{sim(a_i,p_j)}} = 0$. Notice that if any $sim(a_i, p_j) \neq 0$, then the ratio decreases, and the loss increases. After calculating the loss, it is back propagated through the BERT model to fine-tune it for the sentence embedding task.

2. **Reducing the dimensionality of the Sentence BERT embeddings using UMAP**
We would like to create clusters of documents which discuss related topics. BERT sentence embedding creates a 768-dimensional embedding for each sentence. Unfortunately, performing clustering on such high-dimensional data is very ineffective due to the curse of dimensionality. To remedy this issue, the next step in the BERTopic pipeline is to reduce the dimensionality of the sentence embedding to -dimensions so that clustering provides a meaningful result. UMAP is a popular technique which

creates a probabilistic k-nearest neighbor graph in high-dimensional space with Approximate Nearest Neighbor (ANN) methods and subsequently optimizes it with a low-dimensional representation. UMAP focuses on maintaining the structure of local connectivity and performs very well when reducing the dimensionality of high-dimensional data.

3. **Clustering the embeddings using HBDSCAN**
   Now, we would like to cluster the low-dimensional embeddings such that each cluster contains documents which discuss related topics. HBDSCAN is the optimal choice for this scenario since it performs well for highly non-linear embedding spaces and does not require us to know the number of cluster in advance. It also performs outlier detection and labels outlier documents with -1. In our context, that means the content of a given document does not relate to the topics discussed in any cluster of documents.

4. **Constructing a document-term matrix for each cluster using CountVectorizer**
   The next step is to create a document-term matrix for all documents in each cluster. The result is a simple matrix for each cluster where the rows represent each document in the cluster, the columns represent each word in the vocabulary, and the values in the cells are raw counts of the number of times a given word appears in each cluster of documents. This matrix contains intermediate results which are used in the next step. Note that this matrix has shape $C \times V$ where C is the number of clusters and V is the dimensionality of the vocabulary. Additionally, a standard document-terms matrix of shape $D \times V$ is created where D is the number of documents. Later, this is used to give probabilistic assignments to topics for each document.

5. **Creating a more informative document-term matrix using class TF-IDF**
   The document-term matrix created in the last step is not sufficient because it does not account for common, uninformative words. We would like to give more weight to uncommon words since they are more informative. This is why we use class-TF-IDF.

$$\text{c-TF-IDF(t,c)} = \frac{\text{term count of t in cluster c}}{\text{total terms in c}} \log\left(\frac{\text{total number of clusters}}{\text{number of clusters containing t}}\right) \tag{3}$$

   As you can see, the first term simply calculates the term frequency of word w in a cluster of documents c. The second term can be interpreted as a weight for word w. The weight decreases the more word w is used in documents belonging to other clusters. This is intuitive because it means word w is not unique to the topic being discussed in cluster c.

The result of this pipeline is a c-TF-IDF matrix of shape $C \times V$ and a standard term-document matrix of size $D \times V$. Our last step involves calculating a probabilistic distribution over topics for each document. To achieve this, we take a document vector from the term-document matrix ($v_d \in R^V$) and all topic vectors from the c-TF-IDF matrix ($v_t \in R^V$) and calculate a probability distribution using SoftMax and the cosine distance metric. For a document d, we calculate the probability of topic i like so

$$P(t_i|d) = \frac{sim(t_i, d)}{\sum_{j=1}^{T} sim(t_j, d)} \tag{4}$$

If we repeat this for all topics for the document d, we receive a soft assignment of topics which can be used as features for logistic regression. We treat the probability values calculated above as content-related explanatory variables. Since $\sum_{i=1}^{T} P(t_i|d) = 1$ for any d, we do not transform these values prior to regressing them on our explanatory variable.

4.2 Data Cleaning

Once we have the relevant BERT topics, we append this to the existing AGORA dataset. In order to further prepare the dataset for our regression, we first drop many features for the following reasons:

1. **Irrelevance**: Features *Unnamed: 0* or *AGORA ID* are irrelevant to the analysis because they do not include meaningful information.

2. **Text-heavy**: After using the full text of the legislative acts to generate and assign our BERT Topic scores, the text entries are no longer relevant for our analysis.
3. **Collinearity**: To avoid collinearity in our regression, we preemptively drop columns like *Collections* which we anticipate to be correlated with existing explanatory variables (in the case of *Collections*, this would be the *Authority* of the legislation).

The second step in our preprocessing pipeline is to transform the *Authority* field such that we can make better comparisons between the effect of entities. The initial distribution was very long-tailed with many entities having fewer than 3 pieces of legislation in the dataset. *Authority* represents the given entity who is responsible for the piece of legislation. We first filter our observations to only entities in the United States. Then, we further divide entities into federal legislature (US Congress), federal executive (departments of the executive branch and the Executive Office of the President) and state governments. State governments are then grouped into blue, red, and purple given their political tendencies. Finally, we transform this into a one-hot encoding using the federal executive branch as the baseline for comparison.

Third, we remove the existing content related metadata. In the dataset, there was a large amount of content-related metadata with boolean values. There were too many variables to run a meaningful analysis and inclusion of these measures risks collinearity with the BERT topic variables as both are measures of content. Finally, we consolidate the focus of the given legislation (government or private sector) into a single boolean variable and encode the date of most recent activity as an integer representing the number of months since the date of the first piece of legislation in our filtered dataset (Feb 11, 2019). We assume 30 days in a month for this measure.

4.3 Regression

The simplest model using our BERT topic models regresses each BERT Topic x_k on the binary representation of successful enactment (1). We chose to use a logit model in order to bound the predicted values between 0 and 1 and to allow for dynamic marginal effects. The logit model is the basis for all of our regressions. $P(Y=1)$ represents the probability that a proposed piece of legislation is successfully enacted.

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \sum_{k=1}^{K} \beta_k X_k \tag{5}$$

We further expand this model with entity-fixed effects. This accounts for differences between the various state legislatures, the federal legislature, and the federal executive branch. Such effects could include procedural norms or other factors that differ between these entities but not over time. In (2), δ_j represents the entity specific effect for j.

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \sum_{k=1}^{K} \beta_k X_k + \delta_j \tag{6}$$

We also explored the impact of time fixed effects. Broad political environments change over time and as such across entities. Changing rollout of artificial intelligence across the country and other time-related factors may also influence the likelihood a piece of legislation is successfully enacted. In (3), δ_t represents the fixed effect for a given time frame.

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \sum_{k=1}^{K} \beta_k X_k + \alpha_t \tag{7}$$

Then, we add controls and allowing for both time and fixed effects. Note that in this regression we do not allow for any interaction terms. This is explored later on. The control variable that had the most meaningful impact on our predictions is *Private*, a boolean value reflecting whether a given piece of legislation is targeted at private sector entities or public sector entities. Theoretically, the target of legislation may affect its viability to be passed.

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \sum_{k=1}^{K}\beta_k X_k + \delta_j + \alpha_t + \beta_{k+1} Private \tag{8}$$

Our final regression allows for interaction terms. We identified three interaction terms with significant impact on the predicted outcome of a given piece of legislation. All of these terms involve time which in our dataset is *Months after Feb 11, 2019*. Here forward this is simply referred to as *Time*. The first interaction term is the only interaction term between our entity and time fixed effects: *Federal Legislation x Time*. We will discuss the exact implications of this in the results section; however, we felt a strong theoretical basis to include this interaction term as polarization within the US Congress has grown measurably over the past five years. The other interaction terms between entity and time effects had no significant impact and as such are not included in our model. The final two interaction terms present in our model are *Media + Privacy X Time* and *Defense + Education X Time*. The salience of privacy and defense has expanded over this time frame and as such an interaction term allows for a better representation of this empirical reality.

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \sum_{k=1}^{K}\beta_k X_k + \delta_j + \alpha_t + \beta_{k+1} Private + \beta_{k+2} Media*Time + \beta_{k+3} Defense*Time \tag{9}$$

Using a logit regression requires several assumptions including:

1. **No collinearity of predictors**: We address this by dropping known collinear variables; however, the inter-related nature of legislation may mean that certain topics are collinear. We operate under the assumption that such topics are not collinear given their diversity.
2. **Exogeneity**: We assume all predictors to be exogenous, meaning that for any X, $cor(X, u) = 0$. This assumes no reverse causality, measurement errors nor omitted variable bias.
3. **Finite fourth moments**: We know know there are relatively few outliers since all values are bounded by time or binary constraints. Note that BERT Topics are constrained between 0 and 1.

Finally, to evaluate our model, we can compare the Akaike Information Criterion (AIC) scores between models. This is a valuable measure as it reflects the predictive power of the model but penalizes overfitting. The log-likelihood may also be helpful as another measure of fit. Other typical measures that come from the confusion matrix generated by a logit model do not make sense in our context as we are mostly concerned with which variables hold predictive power instead of the specific predictions of the model.

5 Results

Following with the approach outlined in the regression above, the results from each stage of our regression analysis are included below. There were no meaningful lagged dependent variables to include.

	Basic Model	Entity FE	Time FE	Two-Way + Controls	With Interactions
(Intercept)	1.35	4.85**	6.86***	14.72***	6.91*
	(0.90)	(1.68)	(1.30)	(2.30)	(2.74)
Defense + Education	−0.14	1.27	−0.78	1.12	−11.80
	(1.03)	(1.25)	(1.11)	(1.50)	(6.02)
AI + Cybersecurity	0.82	−1.14	1.19	−0.75	0.59
	(1.51)	(2.48)	(1.59)	(2.99)	(3.88)
Energy + Technology	−3.02**	−1.86	−3.30**	−1.79	−1.40
	(1.10)	(1.28)	(1.20)	(1.60)	(1.98)
Media + Privacy	−0.94	−1.93	−0.22	−0.88	−29.90**
	(1.06)	(1.33)	(1.15)	(1.68)	(10.64)
Automated Decision Systems	−2.43*	−3.75**	−2.50*	−3.42*	−2.23
	(1.08)	(1.34)	(1.17)	(1.59)	(1.77)
Foreign Affairs	−4.35**	−2.56	−5.26**	−2.68	−2.24
	(1.38)	(1.37)	(1.77)	(1.79)	(2.07)
Healthcare Services	−2.79*	−3.30*	−2.92*	−3.33	−2.37
	(1.11)	(1.43)	(1.22)	(1.73)	(1.88)
Government + Data Agencies	−1.84	−2.17	−2.23	−3.04	−2.00
	(1.05)	(1.31)	(1.17)	(1.71)	(1.85)
Blue State		−1.99		−2.14	−2.01
		(1.20)		(1.25)	(1.26)
Purple State		−2.05		−2.76*	−2.55
		(1.29)		(1.35)	(1.43)
Red State		−0.43		0.35	0.04
		(1.37)		(1.55)	(1.52)
Federal Legislation		−4.47***		−6.11***	13.51***
		(1.16)		(1.22)	(3.90)
Time			−0.10***	−0.16***	−0.05
			(0.01)	(0.02)	(0.03)
Private Sector Focus				−3.82***	−2.63***
				(0.85)	(0.79)
Media + Privacy:Time					0.51**
					(0.18)
Defense + Education:Time					0.26*
					(0.11)
Federal Legislation:Time					−0.35***
					(0.07)
AIC	493.01	414.89	420.47	289.03	246.77
BIC	528.95	466.82	460.40	348.94	318.66
Log Likelihood	−237.50	−194.45	−200.23	−129.51	−105.38
Deviance	475.01	388.89	400.47	259.03	210.77
Num. obs.	401	401	401	401	401

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Table 1: Regression Results

The Basic Model in Figure 1 represents the regression predicting the survival of a given bill using only its content related factors. This model is largely for illustration as we expect it to suffer from omitted variable bias. The absence of any entity or time effects makes this model ineffective; however, it continues to serve as a meaningful baseline for comparison. The next two models, Entity FE and Time FE allow for the fixed effects individually. In all models, the coefficients represent the change to the log odds of bill survival (standard interpretation of logit coefficients).

We will focus the remainder of our analysis on the final two models which allow for both time and entity fixed effects, control variables, and finally interaction terms. Throughout the progression of the models, the addition of fixed effects tends to increase the intercept as many of the effects have significant negative coefficients. The baseline for our model is a federal executive order in February 2019, and as such, a high predicted probability of success makes sense. As the fixed effects provide greater nuance to deviate from this baseline, the increase in the intercept from the base model to all other models can be understood in the distillation of high predicted probability for the baseline group of bills.

5.1 Fixed Effects + Controls

In the Two-Way + Controls model, the only significant topic is Automated Decision Systems (ADS). This topic is consistently significant with a negative coefficient across the first four models. These models predict that a greater focus on ADS, such as automated loan approvals, hiring decisions and other decision-related automation, decreases a bill's likelihood of survival. The significance of Energy + Technology and Foreign Affairs diminish with the inclusion of both time and entity fixed effects.

The addition of entity fixed effects gives additional nuance to our conclusions. The baseline entity is the federal executive branch and as such all coefficients for entities can be interpreted relative to the likelihood of legislation to pass when introduced via the federal executive branch. This baseline was chosen as these bills face the least democratic scrutiny and therefore would theoretically have the highest probability of being enacted. The results of our entity fixed effects support this hypothesis: all entity fixed effects are either insignificant or negative. Notably, the federal legislature (US Congress) has a strong negative coefficient in both the Entity FE and the Two Way + Controls model. This aligns with the division and procedural politics within the US Congress which limits the ability of this body to easily pass legislation. Our Two-Way + Controls model also detects a significant impact of being a purple state. Partisan division in state government could plausibly decrease the probability of bill passage in a similar way as the US Congress (when compared to the federal executive branch).

Time fixed effects show a strongly significant negative relationship between time and probability of bill passage in both Time FE and Two-Way + Controls. There are several mechanisms which could explain this relationship, several of which we explore in the final model. The addition of the control variable *Private Sector Focus* further improves the model, showing the bills with a focus on the private sector see lower likelihood of enactment compared to equivalent bills for the public sector.

5.2 Interaction Terms

We add three interaction terms to our model, all of which provide further nuance to the impact of time on the survival of a bill. The first two interaction terms are topical, letting time influence the impact of a Media + Privacy or Defense + Education focus on bill survival. Both of these interaction terms have statistical significance, and are supported by empirical evidence. The salience of these topics has grown significantly over the past 5 years and as such the impact of these topics on a given bill could also have changed. Our results support this thinking. The *Media + Privacy:Time* interaction term has a positive coefficient with statistical significance. Notably, although the value of this coefficient is 0.51, time is a continuous variable and as such the impact of this interaction term 10 months after Feb 2019 is already quite large (5.1). To compensate for this large impact, the model produces a negative coefficient for the Media + Privacy topic. This is a better representation of reality as the interaction term allows the impact of media over time to shift positively as concern about media and privacy grows. Such reasoning assumes that growing public salience and democratic consensus would manifest in higher rates of bill enactment. The *Defense + Education:Time* follows a similar interpretation as above given rising geopolitical concerns with the proliferation of artificial intelligence and increasing global tensions.

The final interaction term, *Federal Legislation:Time* is also statistically significant and negative. This predicts that over time, the impact of a bill being proposed in the US Congress (compared to the US Executive Branch), shifts in the negative direction. There is a similar dynamic to the previous interaction terms where we see a large positive coefficient for the Federal Legislature (compared to previous negative coefficients) in order to compensate for this impact. Rising polarization and growing stakes of this legislation are two possible mechanisms which could explain this relationship. Ultimately, the addition of interaction terms increases the predictive power of the model as seen by the shrinking AIC and increasing log likelihood.

To visualize the effect of the described changes, see Figures 1 + 2 below.

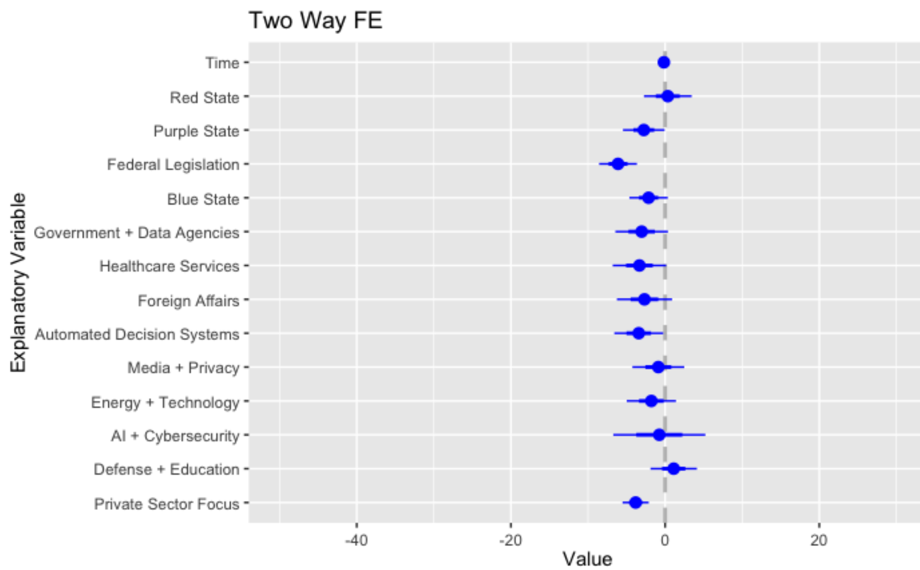


Figure 2: Coefficients in the Two Way Fixed Effects Model (with Control)

Figure 1 represents to coefficients for the Two Way FE model including controls. Figure 2 shows the coefficients after adding interaction terms. Note the shifts in Federal Legislation, Media + Privacy, and Defense + Education as a result of the dynamics described above.

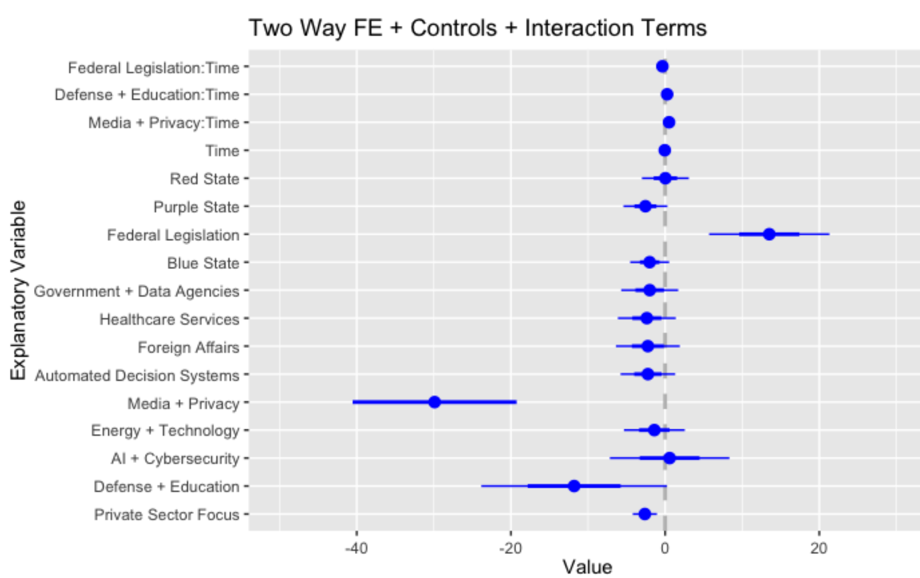


Figure 3: Coefficients after the addition of interaction terms

In summary, our models present several notable takeaways for understanding AI-bill survival in the United States. Our findings highlight the importance of certain topics, like Automated Decision Systems, and show the change in impact of other topics, like Media + Privacy, over time. The significant varied effect of Media + Privacy and Defense + Education over time may allude to changes in the salience or popularity of these themes. The strong effect of Private Sector Focus emphasizes the challenges regulators face when attempting to regulate the private sector as opposed to public entities or citizens. Finally, the fixed effects reiterate the importance of legislative context and partisan politics in passing bills.

6 Conclusion

In this analysis, we aimed to understand the factors that affect the survival of AI-related bills in the United States. We leveraged BERTopic modeling to extract content-related features for each bill and expanded our model to include fixed effects, controls, and interaction terms. We find support for three main takeaways. First, the more homogeneous the respective authority is, the more likely a bill is to pass. This implies that there is significant disagreement between the political fronts on the topic of AI. Second, while the topic of Automated Decision Systems seems to be a particularly heated topic of discussion throughout the observed period, we observe that media, privacy, and defense see more consensus as time goes on. This can be seen as evidence that while opposing views may dominate early discussions, over time, legislators tend to converge on shared understandings of new technologies. Finally, consensus on regulating the private sector seems to be especially difficult to achieve. This could suggest extensive lobbying efforts by private interest groups aiming for unrestricted access to these technologies and their use cases.

Despite these insights, there are some aspects of our analysis that remain shortcomings. Since the dataset relies on legislative acts from the U.S., it is difficult to generalize these insights to other countries. Predicting future cases may also prove challenging if future observations are not properly embedded in their time-related context. Future work could expand the dataset to include other jurisdictions and time frames. Increasing the sample size would allow for further granularity of the entities themselves. Currently, our entity fixed effects are interpreted compared to the baseline of the federal executive branch; however, this is not necessarily a meaningful comparison. Exploring alternative baselines could provide more meaningful analysis, especially as the entities are split into different groups. Furthermore, greater statistical power would allow for more robust modeling of interaction terms.

7 Works Cited

- Adler, E., & Wilkerson, J. (2005). *The Scope and Urgency of Legislation: Reconsidering Bill Success in the House of Representatives.*

- Eidelman, V., Kornilova, A., & Argyle, D. (2018). *How Predictable is Your State? Leveraging Lexical and Contextual Information for Predicting Legislative Floor Action at the State Level.* In E. M. Bender, L. Derczynski, & P. Isabelle (Eds.), *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 145–160). Association for Computational Linguistics. `https://aclanthology.org/C18-1013/`

- Haddii, U. (2025, January). *AI Governance Documents Data.* `https://www.kaggle.com/datasets/umerhaddii/ai-governance-documents-data`

- Nay, J. (2016). *Predicting and understanding law-making with word vectors and an ensemble model* (SSRN Scholarly Paper 2808448). Social Science Research Network. `https://papers.ssrn.com/abstract=2808448`

- Yano, T., Smith, N. A., & Wilkerson, J. D. (2012). *Textual Predictors of Bill Survival in Congressional Committees.* In E. Fosler-Lussier, E. Riloff, & S. Bangalore (Eds.), *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 793–802). Association for Computational Linguistics. `https://aclanthology.org/N12-1097/`