

# Words Matter – Using Machine Learning to Verify How Words Absolve or Condemn Defendants★

Vithor G. F. Bertalan<sup>a</sup> and Evandro E. S. Ruiz<sup>b</sup>

<sup>a</sup>*Polytechnique Montreal, Departement de Genie Informatique et Genie Logiciel. 2500 Chem. de Polytechnique, Montreal, Canada*

<sup>b</sup>*Universidade de São Paulo, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto. Av. Bandeirantes, 3900, Ribeirão Preto, SP – Brazil*

## Abstract

Legal prediction is one of the most critical subfields in Natural Language Processing. The researchers use state-of-the-art machine learning and artificial intelligence methodologies to predict specific judicial facets, such as the judicial outcome. For this research, we have built a web text crawler to extract homicide data cases from Brazilian electronic legal systems. Then, we used word embeddings, processed in several neural networks, to test the ability of our model to predict the outcome of the cases based on their textual characteristics and features, finding that Gated Recurring Units (GRU) showed the best performance. Afterwards, we applied Hierarchical Attention Networks (HAN) to see a sample of the most important words used to absolve or convict defendants. We also analyzed those results to find whether we could track patterns in each of the outcomes.

## Keywords 1

Legal prediction, Attention weights, Natural language processing, Machine learning

## 1. Introduction

Law is among the most text-dependent areas of human activity. Daily new court decisions, appeals, and other written legal instruments are produced by specialized professionals, such as lawyers, judges, defendants, and plaintiffs. All of them have different necessities that intelligent systems could supply. Branting et al. [1] exemplify this potential for using new technologies. In this article, the authors argue that Computer Science has long acted in the automation of legal reasoning, and problem-solving has been an objective of research in Computer Science. Therefore, it is legitimate to consider that Artificial Intelligence (AI) can be used to assist and optimize the daily lives of legal professionals.

With computing resources progressively present in society, today, criminals and law enforcement are increasingly using the opportunities offered by AI. In criminal law, deepfake technologies and algorithmic profiling contribute to existing new types of crime. In criminal procedural law, AI can be used as a law enforcement technology, for example, for predictive policing or cyber agent technology. Machine Learning (ML) is also generally used to improve the legal sector. Ashley and Brüninghaus [2] have also reposted that two long-standing goals in ML and law research are automatic classification of case texts and more precise prediction of case outcomes to make attorneys understand them. In the article by Sula and colleagues [3] we also find arguments that legal professionals benefit significantly from the kind of automation provided by machine learning.

Concerning the data used to feed these AI and ML algorithms, it is worth mentioning that judges have very particular writing styles, as stated in Alarie et al. [4], and often develop particular writing

---

SCIA-2022: 1st International Workshop on Social Communication and Information Activity in Digital Humanities, October 20, 2022, Lviv, Ukraine

Email: vithor.bertalan@polymtl.ca (V. G. F. Bertalan); evandro@usp.br (E. E. S. Ruiz)

ORCID: 0000-0002-1585-7694 (V. G. F. Bertalan); 0000-0002-7434-897X (E. E. S. Ruiz)



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

skills to customize how they present information. Le and his co-authors [5] also characterize legal documents as being very formal, and their structure is sometimes as essential as their readability. Branting et al. [6] state that even if the laws are faithfully represented in the texts, the terms of these same laws are often impossible for a layperson to interpret.

In addition, legal texts have specific properties and structures. The work of Aletras et al. [7] shows us that the formal facts of a court case are the most important predictors of a lawsuit. These descriptive factor texts are essential because, in addition to passages like these, legal texts also have other specific characteristics that differentiate them from other types of narratives. For example, references in a legal text have a specific structure that differs from references in the Public Domain [8]. In [7], we see that textual content and various parts of the case are vital factors influencing the results of the judicial tribunals.

In this project, we applied Hierarchical Attention Networks (HAN) to see a sample of the most important words used to absolve or convict defendants. We also analyzed those results to find whether we could track patterns in each of the outcomes. Our main inspiration for this work is the research performed by Aletras et al. [7], in which textual features were used to predict judicial decisions. To the best of our knowledge, our work is the first to use this methodology, adapted to our final goal of checking the weights of each word in the comprehensive documents to predict judicial decisions in Brazilian Portuguese. The complete research can be found in [9]. In this first section of the paper, we present the introduction and motivation for our research.

In Section 2, we write about similar research that has also inspired our main work. In Section 3, we describe the methodology used in this work, as all as the algorithms chosen. Section 4 presents the discussion about the attention weights of tokens in our judicial texts. In Section 5, we show the results and the discussion conducted on the numbers given in the previous section. Finally, in Section 6, we finish with the conclusion, problems found, and possible future research possibilities.

## 2. Related Works

The main objective of this research lies in the field of sentence prediction. The primary reference to this project is the work of Aletras et al. [7]. In this project, the authors used a dataset of cases from the European Court of Human Rights on cases that violate three articles of its Convention,

namely: Article 3 prohibits torture and inhuman and degrading treatment; Article 6, which protects the right to a fair trial; and Article 8, which provides for the right to respect for private and family life, home and correspondence. From this set of sentences, an equal number of cases that violate (+1) and do not violate (-1) each of the articles were selected and marked. Pre-processing tools and regular expressions for extracting the texts were used. The n-grams were tagged and retrieved in the sections: Procedure, Circumstances, Facts, Relevant Law, Law, and Full Case. These same n-grams were grouped in a vector space model to find the main topics of each article. Classification using Support Vector Machines (SVM) was 78% accurate in predicting topics and circumstances for article 3, 84% in predicting topics and circumstances for article 6, and 78% in predicting topics and circumstances for article 8.

Random Forests' methodology was used in the work of Katz, Bommarito, and Blackman [10] to predict the behavior of the US Supreme Court. The dataset used came from the United States Supreme Court. This set contains 240 variables; among these are: chronological variables, case antecedent variables, justice-specific variables, and outcome variables. Qualitative texts were labeled 'Inverted', 'Affirmed', or 'Other'. The remaining categorical variables were coded with binary or indicator attributes. The authors used the Support Vector Machines and Random Forests methods to classify and predict sentences. The authors pointed to Random Forests as the best algorithm to work on their dataset. Using this methodology, the authors achieved a recall of up to 77% for the 'Affirmed' class, while in binary labeling, they had a recall of 78% for the 'Not Reversed' cases. The authors also studied the weights assigned to terms by the SVM method to find the terms that most impacted each violation of the laws.

Suela and colleagues [3] also carried out an investigation similar to the previous ones in sentence prediction but now involving the areas of Criminal Law and Social or Commercial Law. This group drew on the texts of French Supreme Court decisions (Court de Cassation) and also used the

methodology of SVMs. The decision collection had 131,830 decision documents combined with descriptive metadata. Such standardized metadata consisted of a label for the legal field, a timestamp, the case decision (e.g., forfeiture, rejection, expropriation, etc.), a description of the case, and the cited laws. After pre-processing, the dataset was reduced to 126,865 different court decisions. Each of these documents contained a description of the case and four different types of labels: a) legal area, b) the date of the decision, c) the process itself, and; d) a list of articles and laws cited in the description. The most significant attributes were then selected using hierarchical clustering, and then SVMs were used to classify the dataset. The authors reached a precision of 90.2% of correct answers to classify the area of Law, 96.9% to classify the judicial decision using SVM of 6 classes, and 74.3% using SVM of 7 classes to estimate the date of the process and the decision.

In another study by Gokhale and Fasli [11], the authors developed a learning algorithm for classifying human rights abuses using SVM and logistic regression. They based their work on a domain ontology developed for human rights as background knowledge so that the initial term is extracted to generate labeled data for class training. In the paper by Le and colleagues [5], the authors used index extraction with sentence structure information for Japanese legal documents to assign each token weight, a statistical score showing its importance.

Jurist court verdicts have distinct structures that are particularly suitable for text analysis. In a similar study, Ashley and Brüninghaus [2] also followed this thought. The authors developed a model that extracts information from textual descriptions of facts about the decided cases and applies this information to predict the results of issues. In a similar study, Bertalan [12] used the textual characteristics of tokens to predict court outcomes. However, analysts should not forget that the team led by Nicolas Sannier [13] mentioned that a crucial complexity in analyzing legal texts is that legal provisions are usually interconnected and spread over different texts that cannot be taken in isolation from each other.

In the paper by Sukanya and Priyadarshini [14] we can see several methods that use attention to predict judicial outcomes. All these academic works indicate that the areas of Artificial Intelligence and Natural Language Processing applied to Law is a growing field that offers promising opportunities for research and application in the future. Law is an area of enormous activities in Applied Social Sciences, providing a wide range of applications and generally producing a significant amount of text. Consequently, researchers can produce substantial results with specific applications. The study of computational prediction models for sentences that offer satisfactory results for a sizeable judicial court, such as the São Paulo Justice Court (TJSP), can be of great use to any other court.

### **3. Methods**

This section explains the characterization of domains and the steps necessary to complete the research.

#### **3.1. Data Collection**

We based our work on the data provided by the eSAJ system, the electronic document management system used by the São Paulo Justice Court. The São Paulo Justice Court is the world's largest judicial court, considering the number of legal processes [15]. A crawler was designed and implemented to query and save the documents retrieved.

When one uses eSAJ to query the process's database, the user can select from many fields to exhibit judicial opinions, such as subjects, judges by their names, process numbers, and other fields, to choose the appropriate texts to be retrieved. From the many attributes, have selected the following: Judicial Class, Judicial Subject, Date of availability and Text. Classes are types of judicial documents. For example, repeals and termination of contracts would be judicial classes under Brazilian law. Subjects are the type of judicial case being conducted, such as drug trafficking or femicide. The field Text contains the court decision.

Every process is labeled with its outcomes (absolved or condemned). The average word count for each document is 1.019 words, with the most extensive text accounting for 5.846 words. The crawler can gather the data in a predefined date interval to keep the number of selected cases doable. The initial

date of availability and a final date mentioned in the previous paragraph are initial arguments to capture a restricted corpus. The documents retrieved were available from July 2018 to March 2019. We have collected 1.681 homicide cases, accounting for 844 absolutions and 837 condemnations.

For the labeling process, we hired two Brazilian different criminal lawyers, that were responsible for labeling 40% and 60% of the cases, respectively. That step is necessary due to the complexity of law terms, and since there is no straight information on the outcome of a judicial decision on eSAJ. Hence, the lawyers read the whole content of the decision so as to define whether the defendant had been absolved or convicted.

### **3.2. Data Transformation**

Instead of using the words in the documents, the words were transformed into an embedded form. Turian, Ratinov, and Bengio [16] define word embeddings as vectors composed of real numbers distributed over an interdimensional space induced by semi-supervised learning. Since machine learning algorithms work with numbers, not words, word embeddings are a very effective alternative to transforming words into mathematical values. The reasoning for word embeddings is to calculate the similarities between two vectors by cosine similarity. Each vector dimension represents a characteristic intending to capture a word's semantic, in other words, its synthetic or morphological proprieties in a distributed way.

GloVe [17] was the chosen methodology used to construct the embedded words used in the project. GloVe learns word vectors such that their dot product equals the logarithm of the words' probability of co-occurrence. Rather than using a window to define local context, GloVe constructs an explicit word-context or word co-occurrence matrix using statistics across the whole text corpus. It can combine local and global representations of a term by mixing the features of two model families: the global matrix factorization and local context window methods. We have used the pre-trained GloVe corpora developed by the team led by Sandra Aluisio [18] for Brazilian Portuguese.

### **3.3. Machine Learning Processing**

We used five different neural network algorithms to process the dataset. These are Multilayer Perceptron (MLP), Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM), Gated Recurring Unit (GRU), and Hierarchical Attention Networks (HAN). We also applied a 10-fold cross-validation procedure in all five algorithms. Every file was embedded using Brazilian Portuguese GloVe word vectors with 600 embedding dimensions. Both the learning rate and the loss function define the convergence criteria. We used those word embeddings on GloVe, processed in several neural networks, to test the ability of the model to predict the outcome of the cases based on their textual characteristics and features. The learning rate was lowered by 0.2 every three epochs. Concerning the loss function, the algorithm stops if, after five epochs, it does not decrease by at least 0.001. Results were evaluated through the standard quality evaluation measures: accuracy, precision, recall, and F-measure. In our tests, we have tried several different combinations of hyperparameters, to see which ones offered the best results, and to find the optimal combination for each of the algorithms.

### **3.4. Using Hierarchical Attention Networks**

Although we used many different neural networks to evaluate the dataset, the research's core revolves around utilizing Hierarchical Attention Networks (HANs). Yang et al. [19] describe a HAN as a neural network architecture that highlights the importance of individual words, or phrases, in building the representation of a document. This model emphasizes the most important sequences of terms that affect the classification of a document. In theory, it is known that not all terms are equally crucial for the classification of a text and that not all sentences represent the same meaning either.

The result of processing HAN networks over a text is the association of an attention coefficient for each term, indicating the importance of that term in its sentence. Still, considering the article by Yang et al. [19], we can see the application of a HAN in Figure 1. Note that sentence 1 ("*The woman was*

*present at the crime scene*") is more critical than sentence 2 for the overall text label (the distinguishable shades of gray exemplify this difference).

This sentence was highlighted in the HAN network because the network ranked this sentence among the most important in this text - that is, this sentence received a relatively greater weight of attention than the others. Also, note that in this sentence, some words are marked in red. These words constitute essential terms in these sentences, as they have the highest attention weights of the words.

The woman was present at the crime scene.

Although she was carrying no weapons, her hands

were filled with blood and gun powder.

**Figure 1:** Example of word attention weights generated by an Hierarchical Attention Networks(HAN) [19].

## 4. Results

In this section, we describe the numerical results and the textual analysis performed on the attention weights of the HANs.

### 4.1. Numerical Evaluation of Results

We used four distinct networks to evaluate the numerical results: a Multilayer Perceptron network, an RNN (recurrent neural network), an LSTM type network, Long-short term memory, and, finally, a GRN, Gated Recurrent Network.

For the Multilayer Perceptron network, we adopted the following configuration: the architecture of 3 hidden layers, with 512 X 512 X 250 neurons, respectively, executed in 25 epochs. For the RNNs, we use a hidden layer architecture with 128 units, with a dropout probability of 0.5 for the hidden layer and a dropout probability of 0.2 for the inputs. The sigmoid function was used as the activation function, and the binary cross-entropy function as the loss function performed in 25 epochs.

For the LSTM networks, a hidden layer architecture with 128 units was also used, with a dropout probability of 0.2 for the hidden layer and a dropout probability of 0.2 for the inputs. For the tests using the GRN network, we also used a 1-hidden layer architecture with 128 units and a probability equal to 0.2 for both a 0.2 dropout in the hidden layer and the inputs. In contrast, all HANs were used with 600 dimensions of embedding, with a word and phrase encoder. The results can be found in Table 1.

**Table 1**

Evaluation of different Neural Network approaches.

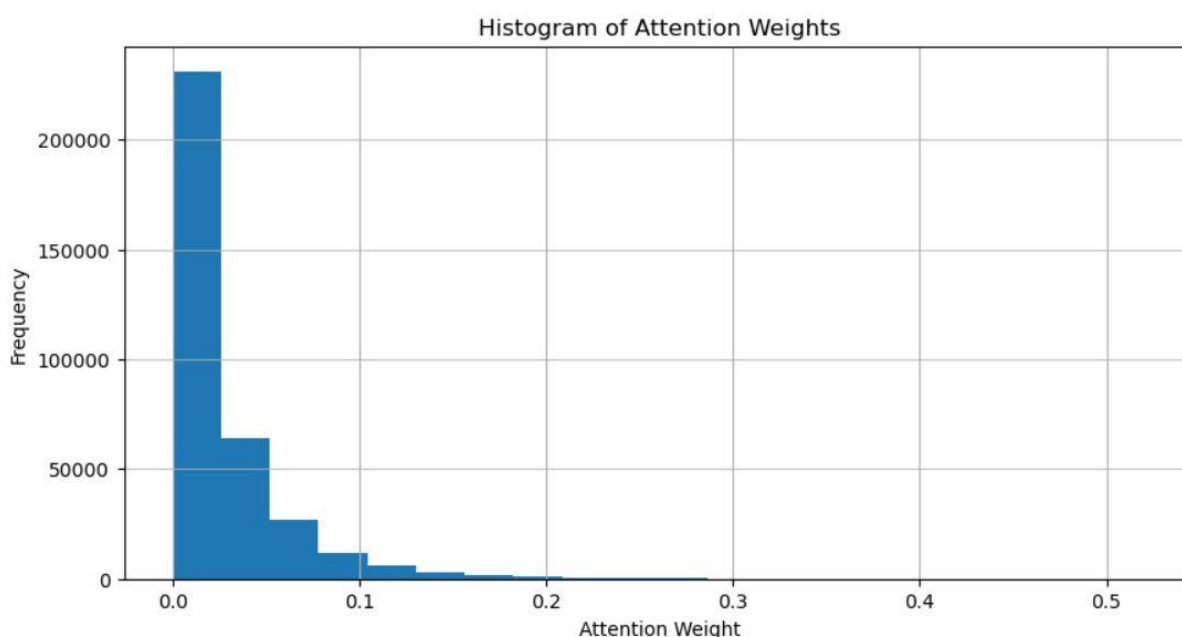
Neural Network	Precision	Recall	F-Score	Accuracy
MLP	0.98	0.98	0.98	0.98
RNN	0.86	0.88	0.86	0.85
LSTM	0.98	0.98	0.98	0.98
GRU	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>
HAN	0.96	0.98	0.97	0.98

Our results are similar to Chung et al. [20]. Both have shown that the GRU is faster than other neural network architectures while showing comparable accuracies. Chung et al. also mention that the choice of the type of network may depend heavily on the dataset and corresponding task. Considering this particular dataset, our research shows that GRU is the best choice.

## 4.2. Computation of Attention Weights

All attention weights were calculated for each word in the data sets. Homonymous words appear several times in the dataset corresponding to their different meanings. Note that the same word can have different weights of attention for each class involved, acquittal or condemnation.

We have recorded 248,460 single words for the acquittal texts in alleged murders and 466,461 single word cards for convictions for murder. For each of the tokens, their attention weight was calculated.



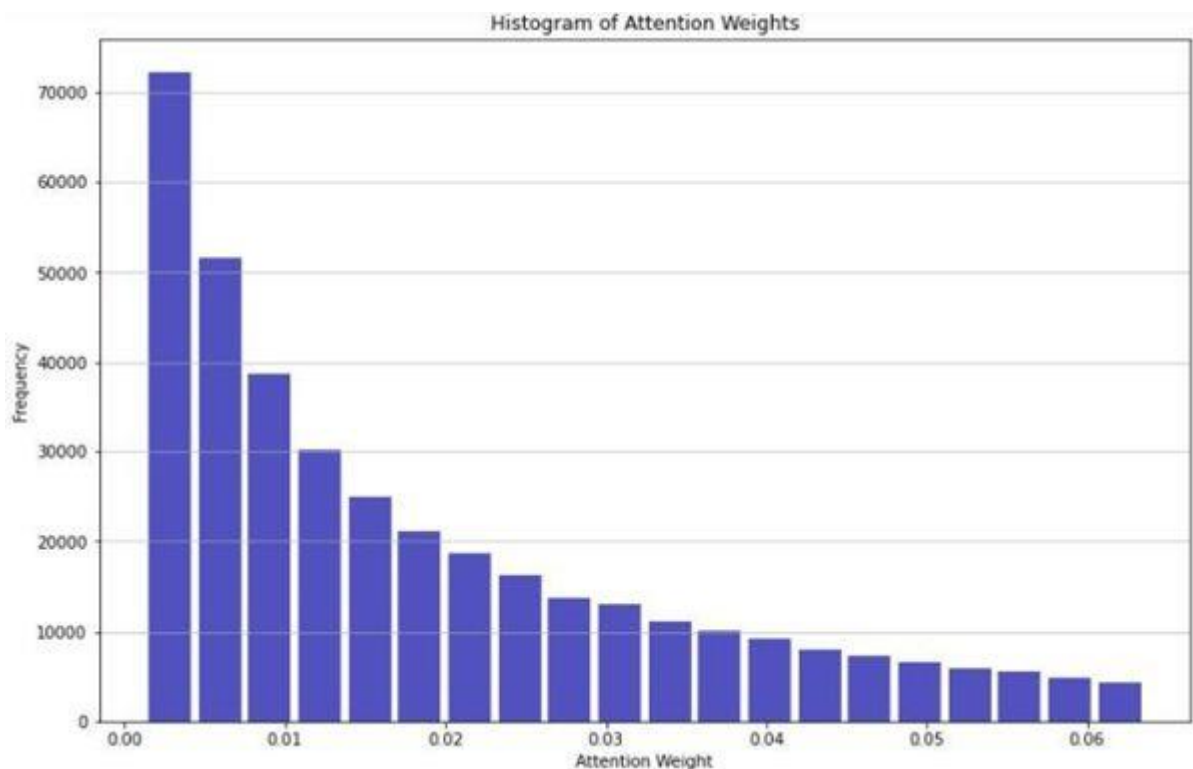
**Figure 2:** Histogram of word attention weights for the homicide condemnations.

A histogram of word attention weights may be seen in Figure 2. The slope of this histogram may be explained through Zipf's Law. By this law, considering a relevance metric, it is possible to infer that the frequency of any word is inversely proportional to its rank. Accordingly, the relevant terms account for a small proportion of our dataset.

The graphic in Figure 3 does not show either the top 10% or the low 10% that were removed, and it shows only the 80% middle words remaining. We can see that Zipf's Law continues to explain the behavior of the terms. Although Figure 3 refers to the condemnation documents, the same behavior was observed in the absolution documents.

After the classification had been performed, the words in each dataset were ordered by their attention weights. Each word had a unique value, ranging from 0 (a word with no importance for document classification) to 1 (for a word with maximum importance for document classification). It is worth mentioning that a word might have different attention weights in distinct sentences.

As an example, consider these two sentences: *The defendant robbed a bank* and; *The defendant did not participate in the robbery, because it was going to a blood bank*. Both have the word *bank*, but in two very different contexts. In the first sentence, the word would be a vital contributor to the condemnation, while the same word would contribute to the absolution in the second sentence. Therefore, words have appeared more than once in our final calculations, with different attention weights. The list with the top 50 words for each of the four outcomes is listed in Table 2.

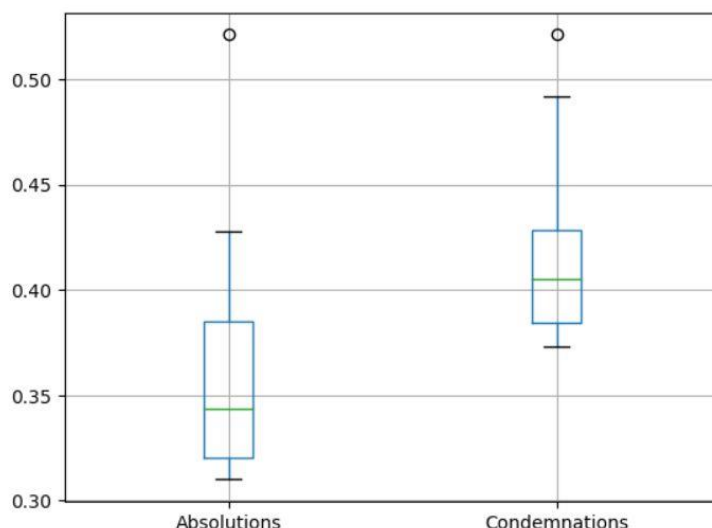


**Figure 3:** Histogram of 80% word attention weights on the middle interval.

## 5. Discussion and Conclusions

The first item we can analyze from the results is the difference in the number of words of absolutions and condemnations: 248,460 unique word tokens for absolutions and 466,461 unique tokens for condemnations, an 86% increase between the two groups. Therefore we infer that judges, as well as their clerks, tend to write more in condemnation texts and dispatches – or, at least, they try to use a vocabulary more refined and distinctive in those cases. That can reflect a more significant amount of the established jargon in condemnation cases, indicating that those types of outcomes have particular words, or that Law professionals tend to write more and more differently in those types of judicial cases.

Also, when we analyze the dimension of the weights, we see that the condemnation tokens have a more substantial weight on the outcome of the case. Even though the top two tokens (coincidentally, "bo" for both outcomes have the same weight). However, when we look down on the list, we see that the condemnation words keep on having a high weight. The 50th weights 0.373, while the 50th in the absolution list weights 0.310. From that, we can observe that the unique tokens in the condemnation list have a higher impact on the outcomes than absolutions. This fact matches the previous observation, indicating that the use of a more advanced vocabulary for those cases significantly affects the text as a whole. We can see this difference graphically in Figure 4.



**Figure 4:** Boxplot of word attention weights.

Looking directly at the meaning of words, we can also see patterns in the lists of both outcomes. In the absolution list, we can read many words relating to the context of the defendant, such as *bem* (good), *origem* (origin), *infância* (childhood), *social* (social), *tititi* (gossip), *cor* (color), *mãe* (mother), *anos* (years), and *soubessem* (if they knew). Although further analysis is required, we can conclude that many absolution texts consider the socio-cultural aspects of the defendant to substantiate the decision.

On the other hand, condemnation unique tokens contain several references to violent terms, such as *homicídio* (homicide), *qualificado* (aggravated), *disparos* (gun shots), *golpes* (blows), *socos* (punches), *lesões* (lesions) and *colisão* (collision). Also, several terms relating to the judicial process, such as *infração* (infraction), *penal* (criminal), such as *sentença* (sentence), *acusação* (accusation) and *comarca* (county). We can see by those terms that condemnation texts tend to heavily emphasize the nature of the crime and the judicial process to substantiate the criminal penalty.

It is helpful to mention that a word might have different attention weights in distinct sentences. As a short example, the sentence *The defendant robbed a bank* and the sentence *The defendant did not participate in the robbery, because it was going to a blood bank* both have the word *bank*, but in very different contexts. In the first sentence, the word would be a vital contributor to the condemnation, while it would contribute to the absolution in the second sentence.

## 6. Conclusions

By analyzing the attention weights of each outcome, we could see that words do have substantial effects on the meaning of the text. All the mathematical coefficients for the attention weights are shown in Table 2. A HAN analysis offers a significant advantage when we compare our results to a single word count. When we use attention neural networks, we can mathematically capture the impact of each word in a single sentence, and each sentence in the overall text.

However, we have to consider such a technique's social and technical issues. As for technical issues, as described by Surden [21], there are certain well-known limitations to applying AI in Law. First, the model will be helpful only if a future case class has standard features pertinent to previous analyzed topics in the training set. It will only be helpful if the model has the same properties as the previous case. Therefore, the model does not consider the subtle changes in judicial thinking over time, except when these changes represent a significant volume of training data. It will only consider subtle changes in judicial thinking. The authors also present an example: not every law firm has several cases that are similar enough to each other that the previous case has elements helpful in predicting future outcomes. Therefore, one might infer that only large law firms will possess the financial and technological power to develop such models.

Regarding the social issues, Katz, Bommarito, and Blackman [10] state that qualitatively oriented legal experts tend to suggest improvements in the model based on anecdotes or their untested mental



model rather than reliable facts and factual data. The authors suggest that to support a case from the future applicability of a model, it must consistently outperform a baseline comparison. This prerequisite is necessary not only for scientific purposes but also to gain lawyers' trust for the model. We can also see potential strategies to address gaps in social problems with Artificial Intelligence in Santoni and Mecacci [22].

A future path of research could also consider the sentences, not just the unique tokens, to perform a contextual analysis of the meaning of the texts. This way, the most important sentences for each outcome could also be considered and combined with the analysis we did in this paper to determine which complex expressions have the highest weight in each of the outcomes. Another possibility is to extend the research to the legal systems of other countries or languages to identify if our findings are consistent with those different scenarios or if each language has its behavior pattern in its legal texts.

**Table 2**

Word attention weights for the homicide dataset, and their English translations.

Homicide Absolutions			Homicide Condemnations		
Position	Token	Weight	Position	Token	Weight
1	bo ( <i>incident report</i> )	0.521	1	bo ( <i>incident report</i> )	0.521
2	mogi ( <i>Brazilian city</i> )	0.428	2	cristina ( <i>Brazilian name</i> )	0.492
3	comarca ( <i>county</i> )	0.417	3	horário ( <i>time</i> )	0.479
4	estado ( <i>state</i> )	0.416	4	infração ( <i>infraction</i> )	0.464
5	santos ( <i>Brazilian city</i> )	0.414	5	penal ( <i>criminal</i> )	0.456
6	justiça ( <i>justice</i> )	0.413	6	cf ( <i>Federal Constitution</i> )	0.452
7	bem ( <i>good</i> )	0.411	7	regime ( <i>regime</i> )	0.445
8	sala ( <i>room</i> )	0.407	8	xavier ( <i>Brazilian name</i> )	0.444
9	cep ( <i>postal code</i> )	0.407	9	homicídio ( <i>homicide</i> )	0.442
10	competência ( <i>competence</i> )	0.403	10	qualificado ( <i>aggravated</i> )	0.442
11	antes ( <i>before</i> )	0.389	11	disparos ( <i>gun shots</i> )	0.440
12	volta ( <i>return</i> )	0.387	12	sant ( <i>unknown token</i> )	0.438
13	origem ( <i>origin</i> )	0.387	13	exposto ( <i>exposed</i> )	0.430
14	infância ( <i>childhood</i> )	0.379	14	provisório ( <i>provisory</i> )	0.424
15	social ( <i>social</i> )	0.373	15	mediante ( <i>through</i> )	0.422
16	porque ( <i>why</i> )	0.369	16	philipe ( <i>Brazilian name</i> )	0.421
17	júri ( <i>jury</i> )	0.363	17	sentença ( <i>sentence</i> )	0.419
18	principal ( <i>main</i> )	0.352	18	toledo ( <i>Brazilian name</i> )	0.417
19	júri ( <i>jury</i> )	0.352	19	osmarina ( <i>Brazilian name</i> )	0.416
20	altura ( <i>height</i> )	0.349	20	juízo ( <i>in court</i> )	0.415
21	placas ( <i>signs</i> )	0.349	21	ip ( <i>police investigation</i> )	0.415
22	nunes ( <i>Brazilian name</i> )	0.348	22	narra ( <i>tells</i> )	0.413
23	p ( <i>page</i> )	0.348	23	golpes ( <i>blows</i> )	0.409
24	machado ( <i>Brazilian name</i> )	0.346	24	justiça ( <i>justice</i> )	0.409
25	porte ( <i>weapon carry</i> )	0.344	25	sala ( <i>room</i> )	0.407
26	agnaldo ( <i>Brazilian name</i> )	0.343	26	acusação ( <i>accusation</i> )	0.403
27	sp ( <i>Brazilian state</i> )	0.338	27	sentença ( <i>sentence</i> )	0.401
28	anos ( <i>years</i> )	0.335	28	marta ( <i>Brazilian name</i> )	0.400
29	regina ( <i>Brazilian name</i> )	0.332	29	estado ( <i>state</i> )	0.398
30	tititi ( <i>gossip</i> )	0.328	30	silva ( <i>Brazilian name</i> )	0.394
31	permitido ( <i>allowed</i> )	0.327	31	estado ( <i>state</i> )	0.393
32	cor ( <i>color</i> )	0.326	32	sassolli ( <i>Brazilian name</i> )	0.391
33	mãe ( <i>mother</i> )	0.325	33	prisão ( <i>prison</i> )	0.389
34	josé ( <i>Brazilian name</i> )	0.322	34	rua ( <i>street</i> )	0.387
35	instrução ( <i>instruction</i> )	0.322	35	sentença ( <i>sentence</i> )	0.386
36	cento ( <i>cent</i> )	0.322	36	justiça ( <i>justice</i> )	0.385
37	comum ( <i>common</i> )	0.321	37	socos ( <i>punches</i> )	0.385
38	réu ( <i>defendant</i> )	0.320	38	análise ( <i>analysis</i> )	0.384
39	cosmópolis ( <i>Brazilian city</i> )	0.319	39	flores ( <i>flowers</i> )	0.382

40	estado ( <i>state</i> )	0.319	40	estrita ( <i>strict</i> )	0.382
41	saído ( <i>gone</i> )	0.318	41	mínimo ( <i>minimum</i> )	0.382
42	soubessem ( <i>if they knew</i> )	0.317	42	competência ( <i>competence</i> )	0.381
43	ação ( <i>action</i> )	0.317	43	lesões ( <i>lesions</i> )	0.379
44	tribunal ( <i>court</i> )	0.314	44	infância ( <i>childhood</i> )	0.379
45	pública ( <i>public</i> )	0.313	45	artigos ( <i>articles</i> )	0.377
46	ordinário ( <i>ordinary</i> )	0.313	46	colisão ( <i>collision</i> )	0.377
47	todos ( <i>all</i> )	0.312	47	regime ( <i>regime</i> )	0.376
48	central ( <i>central</i> )	0.311	48	causaram ( <i>caused</i> )	0.376
49	sessenta ( <i>sixty</i> )	0.311	49	comarca ( <i>county</i> )	0.374
50	júri ( <i>jury</i> )	0.310	50	dinheiro ( <i>money</i> )	0.373

## 7. References

- [1] L. K. Branting, A. Yeh, B. Weiss, E. Merkhofer, B. Brown, Inducing Predictive Models for Decision Support in Administrative Adjudication, in: U. Pagallo, M. Palmirani, P. Casanovas, G. Sartor, S. Villata (Eds.), *AI Approaches to the Complexity of Legal Systems*, Springer International Publishing, Cham, 2018, pp. 465–477.
- [2] K. D. Ashley, S. Brüninghaus, Automatically classifying case texts and predicting outcomes, *Artificial Intelligence and Law* 17 (2009) 125–165. doi:10.1007/s10506-009-9077-9.
- [3] O.-M. Sulea, M. Zampieri, S. Malmasi, M. Vela, L. P. Dinu, J. van Genabith, Exploring the Use of Text Classification in the Legal Domain, in: *Proceedings of the 2nd Workshop on Automated Semantic Analysis of Information in Legal Texts (ASAIL)*, 2017. URL: <http://arxiv.org/abs/1710.09306>. arXiv:1710.09306.
- [4] B. Alarie, A. Niblett, A. Yoon, How Artificial Intelligence Will Affect the Practice of Law, in: *Artificial Intelligence, Technology and the Future of Law*, 2017, pp. 1–16. URL: [https://papers.ssrn.com/sol3/papers.cfm?abstract={\\_}id=3066816](https://papers.ssrn.com/sol3/papers.cfm?abstract={_}id=3066816).
- [5] T. T. N. Le, K. Shirai, M. L. Nguyen, A. Shimazu, Extracting indices from Japanese legal documents, *Artificial Intelligence and Law* 23 (2015) 315–344. doi:10.1007/s10506-015-9168-8.
- [6] L. K. Branting, A. Yeh, B. Weiss, E. Merkhofer, B. Brown, Inducing predictive models for decision support in administrative adjudication, in: *AI Approaches to the Complexity of Legal Systems*, Springer, 2015, pp. 465–477.
- [7] N. Aletras, D. Tsarapatsanis, D. Preoțiuc-Pietro, V. Lampos, Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective, *PeerJ Computer Science* 2 (2016) e93. URL: <https://peerj.com/articles/cs-93>. doi:10.7717/peerj-cs.93.
- [8] O. T. Tran, B. X. Ngo, M. L. Nguyen, A. Shimazu, Automated reference resolution in legal texts, *Artificial Intelligence and Law* 22 (2014) 29–60. doi:10.1007/s10506-013-9149-8.
- [9] V. G. F. Bertalan, Using natural language processing methods to predict judicial outcomes, Ph.D. thesis, Universidade de São Paulo, 2020.
- [10] D. M. Katz, M. J. Bommarito II, J. Blackman, A general approach for predicting the behavior of the Supreme Court of the United States, *PLoS one* 12 (2017) e0174698.
- [11] R. Gokhale, M. Fasli, Deploying A Co-training Algorithm to Classify Human-Rights Abuses, in: *2017 International Conference on the Frontiers and Advances in Data Science (FADS)*, 2017, pp. 108–113.

- [12] V. G. F. Bertalan, E. E. S. Ruiz, Predicting judicial outcomes in the brazilian legal system using textual features., in: DHandNLP@ PROPOR, 2020, pp. 22–32.
- [13] N. Sannier, M. Adedjouma, M. Sabetzadeh, L. Briand, An automated framework for detection and resolution of cross references in legal texts, *Requirements Engineering* 22 (2017) 215–237. doi:10.1007/s00766-015-0241-3.
- [14] G. Sukanya, J. Priyadarshini, A meta analysis of attention models on legal judgment prediction system, *International Journal of Advanced Computer Science and Applications* 12 (2021).
- [15] Tribunal de Justica de Sao Paulo – Quem Somos, 2022. URL: <https://www.tjsp.jus.br/QuemSomos>.
- [16] J. Turian, L. Ratinov, Y. Bengio, Word representations: a simple and general method for semi-supervised learning, in: *Proceedings of the 48th annual meeting of the association for computational linguistics*, Association for Computational Linguistics, 2010, pp. 384–394.
- [17] J. Pennington, R. Socher, C. D. Manning, GloVe: Global Vectors for Word Representation, in: *Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543. URL: <http://www.aclweb.org/anthology/D14-1162>.
- [18] N. Hartmann, E. Fonseca, C. Shulby, M. Treviso, J. da Silva, S. Aluisio, Portuguese Word Embeddings: Evaluating on Word Analogies and Natural Language Tasks, arXiv preprint arXiv:1708.06025 (2017). URL: <https://arxiv.org/abs/1708.06025>.
- [19] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, E. Hovy, Hierarchical Attention Networks for Document Classification, in: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, San Diego, California, 2016, pp. 1480–1489. URL: <https://www.aclweb.org/anthology/N16-1174>. doi:10.18653/v1/N16-1174.
- [20] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv preprint arXiv:1412.3555 (2014).
- [21] H. Surden, *Artificial Intelligence and Law*, *Washington Law Review* 89 (2014) 87–116.
- [22] F. Santoni de Sio, G. Mecacci, Four responsibility gaps with artificial intelligence: Why they matter and how to address them, *Philosophy & Technology* (2021) 1–28.