Predicting the Duration of Judicial Cases Using Hybrid Systems Based on Language Models

Amina BOUHOUCHE[1], Saliha YASSINE[2], Mustapha ESGHIR[3], Mohammed ERRACHID[4]
Faculty of Sciences LabMIA-SI, Mohammed V University in Rabat, Rabat, Morocco[1, 2, 3]
Regional Center for Education and Training Professions (CRMEF), Rabat, Morocco[4]

***Abstract*—Recent technological developments in the field of Natural Language Processing (NLP), notably due to Transformer architectures and language models, have made it possible to tackle aspects that were previously inaccessible with traditional tools. The present study addresses the issue of predicting legal case durations using Arabic judicial data. For this task, hybrid systems based on language models were implemented. The Arabic_LegalBERT model, derived from AraBERT and specialized through additional pre-training on an Arabic legal corpus, was proposed to generate representations that were integrated into the downstream steps of the approach. Two methods were adopted for predicting the processing time of a new case: The first followed a framework combining automatic classification with statistical correspondence, while the second relied on cosine similarity combined with empirical statistics. The results obtained with the classification approach are particularly promising, with a small improvement for the system based on the specialized model. For the similarity-based approach, the results are also promising, with a clear distinction observed when evaluating each type individually, indicating that types with a higher number of cases generally perform better than those with fewer cases.**

Keywords*—Language model; judicial case durations; legal domain; Arabic legal corpus

I. Introduction

The application of NLP and machine learning tools in the legal field is a rapidly growing area of research. Technological advances have enabled researchers to better leverage the large volume of legal documents and address various tasks specific to this domain using advanced methods. These include transformers, which have revolutionized traditional NLP approaches thanks to the attention mechanism [1]. Their architecture has served as the basis for more advanced language models capable of specializing in particular domains and specific languages. This specialization can be achieved by training a language model from scratch, thereby treating the domain-specific language as a language in its own right [2] [3], or by leveraging pre-trained general models and adapting them to the target domain through additional pre-training on specialized corpora.

The present study addresses the prediction of judicial case durations, a task specific to the legal domain, whose main impact lies in the organizational optimization it can provide within the judicial system. Furthermore, the study is based on legal data in Arabic, a language that is both rich and complex, and that exhibits notable differences compared to Western languages, requiring the use of tools adapted to its specific characteristics. Despite its importance, this issue remains largely underexplored in existing research, and even more so in the context of the Arabic language. Thus, addressing this issue involves multiple dimensions, including the exploration of NLP tools, the use of language models specialized in Arabic and the legal field, and empirical statistical techniques. Indeed, the study first focuses on the specialization of the Arabic_LegalBERT model on an Arabic legal corpus, through additional pre-training applied to the AraBERT model [4], which itself is based on BERT (Bidirectional Encoder Representations from Transformers) and specialized in the Arabic language. Next, two hybrid systems are adopted to predict the duration of a judicial case. The first consists of fine-tuning Arabic_LegalBERT with a classification approach combined with statistical calculation, and the results are compared with those obtained using AraBERT in the same approach. The second is based on a cosine similarity approach combined with empirical statistics. For this similarity approach, the weighted average of the most similar cases identified was compared to the average of the majority type cases among those identified. The impact of the data on the results was also evaluated, indicating that a larger corpus would allow for further exploration of this specific task.

This study focused on seven sections: Section II presents a review of previous work, particularly on language models specialized in the legal field. Section III addresses the issue of judicial delays, their causes, implications, and challenges. Section IV is dedicated to describing the methodology used, including the data used, the adapted model, and the experimentation carried out. Section V is devoted to describing the results obtained, while Section VI discusses them with a focus on the limitations encountered. The conclusion presented in Section VII summarizes the key points of the study and opens the way for further improvement.

II. Related Works

The legal domain is characterized by the diversity of its sources and content, and the vast amount of available legal data makes it necessary to rely on advanced tools to accomplish a given task. Previous research has addressed the legal field using traditional machine learning models [5] [6] [7], then evolved towards deep models [8] [9] [10]. At the same time, researchers have benefited from the evolution of text representation methods, particularly contextual representations based on Transformer architectures, integrated into recent language models.

Language models, such as BERT [11], have demonstrated particularly interesting capabilities, prompting many researchers to develop derivative variants. Some models have been designed for specific domains, such as science with SciEBERT [12], biomedicine with BioBERT [13], and finance with FinBERT [14]. Other specialized models have been developed for specific languages, such as the Italian model AlBERTo [15], the French model CamemBERT [16], and the Spanish model BETO [17]. Regarding models specialized in the Arabic language, it is essential to mention the AraBERT model, which has been pre-trained on a vast Arabic language corpus and achieves cutting-edge performance in several NLP tasks [4]. There are also the ARBERT and MARBERT models, pre-trained on a very large and varied Arabic corpus, with a particular advantage for MARBERT, which covers not only Modern Standard Arabic (MSA) but also Arabic dialects [18].

The legal domain, which is the focus of this study, has also been the subject of research leveraging the potential of language models. Reference [19] is one of the first studies to apply the BERT model to the legal domain in English, using a hierarchical version of BERT. In a subsequent study, a LEGAL-BERT model was trained on English legal data to handle legal tasks [2]. The authors of this work evaluated two strategic approaches: fine-tuning and training from scratch, and demonstrated that the models resulting from both approaches outperform the original BERT model. Subsequently, new perspectives opened and several derivatives of the BERT model specializing in the legal domain appeared. In [20], the authors focused on the knowledge distillation technique adopted in the DistilBERT model [21] to train its own LegalDB model on a legal corpus. In addition, other derivatives of the BERT model have been specialized in the legal field, in languages other than English. These include the JuriBERT model [22], specifically designed to process French legal language, the Legal Spanish Language Model (MEL), derived from the XLM-RoBERTa-large model, refined from Spanish legal corpora [23], and the ITALIAN-LEGAL-BERT models, pre-trained on Italian legal language and outperforming general Italian models in classification, named entity recognition, and semantic similarity tasks [24]. For models specializing in the legal field in Arabic, it is worth mentioning the AraLegal-BERT model [3], which was trained from scratch on a large corpus of purely legal documents. After the training phase, the model was refined to handle three natural language understanding (NLU) tasks. Based on the results obtained, the authors emphasize the positive impact of domain-specific pretraining over general-purpose models for certain NLU tasks [3]. The study [25] also looked at legal prediction, focusing mainly on Arab commercial cases. The authors used three language models (LLaMA-7b, JAIS-13b, and GPT-3.5-turbo) and three training paradigms (zero-shot, one-shot, and fine-tuning). They highlighted the potential of the GPT-3.5 model, while emphasizing the difficulties encountered in predicting decisions and the limitations related to the quality and availability of Arabic judicial data.

III. Judicial Case Durations

In this study, we focused on judicial case duration, a fundamental element of the judicial system and a key indicator of its efficiency. The contribution of our work lies in the analysis of a dimension that is rarely addressed in current research. Achieving a reasonable timeframe in legal proceedings is a shared objective, not only of the various stakeholders in the judicial system, but also of litigants and States. The 2011 Moroccan Constitution stipulates that 'Everyone has the right to a fair trial and to a judgment rendered within a reasonable time.' However, the length of proceedings is affected by the combined effect of multiple factors. This time limit may vary from one case to another depending on multiple causes that are not solely attributable to the judicial institution. Certainly, delays may be linked to internal dysfunctions such as court overload and insufficient human and material resources, however, there are also external factors that directly impact the length of legal proceedings. Case complexity is a major factor, as complex cases require in-depth analysis and greater allocation of effort and resources. In addition, witness attitude and the use of judicial expertise can further influence the duration of proceedings. Ultimately, judicial case durations can be understood as a multidimensional variable, explained by a combination of institutional, procedural and behavioral imperatives. The deployment of intelligent systems capable of estimating the duration of legal proceedings could optimize the management of justice. This system could enable professionals to optimize the organization of the cases they handle, evaluate their number and criticality, anticipate potential delays with respect to deadlines, and assess their legitimacy.

IV. Methodology

Given that the task under study involves estimating the duration of legal cases, this work contributes by combining the advantages of language models with classification and similarity-based approaches, along with empirical statistics, to tackle a task that integrates both textual data and the temporal dimension. Traditional machine learning models have proven effective for various tasks but remain unable to capture the semantic and contextual aspects of the data. For this reason, this study relies on representations generated by language models, which are subsequently utilized in downstream modules. The model developed in this work is Arabic_LegalBERT, and the duration of legal proceedings is estimated by using two systems: one based on classification and the other on cosine similarity, combined with empirical statistics.

A. Dataset

Data relating to court cases is very scarce, particularly for cases in Arabic. Furthermore, the available data sometimes lacks information that is relevant for a specific analysis. In the context of this work, some cases lack the dates necessary for calculating case durations. For this study, a set of court cases in Arabic was collected manually. The cases collected are of different types and can be divided into four categories: payment cases, rental disputes such as rent reviews and evictions, disputes relating to traffic accidents, and other disputes, including claims for compensation, claims for cessation of harm, and others.

The empirical statistics for the four types of cases are summarized in Table I. The total number of cases is 329, distributed unevenly among the four types studied. The

statistics show that the durations are highly dispersed, particularly for cases relating to traffic accidents and miscellaneous cases. This is reflected in the respective standard deviations of 129 and 195 days, as well as in the corresponding ranges of 470 and 696 days. Conversely, payment cases and rental disputes appear relatively stable, with lower standard deviations of 63 and 81days and reduced ranges of 252 and 289 days. This heterogeneous dispersion can be explained by the nature of the cases. Payment and rental cases, for example, are relatively homogeneous due to the standardized and well-established procedures that generally characterize these types of cases. On the other hand, traffic accident cases and miscellaneous cases are somewhat unique. Depending on the case, they may require technical or medical expertise, the presence of several stakeholders and other factors. All these factors contribute to the fact that the timeframes involved are more dispersed.

B. Model

The model proposed in this study designated Arabic_LegalBERT is based on AraBERT, specifically the Bert-base-arabert version available on Hugging Face Hub. To this end, an additional pre-training phase was added to AraBERT to specialize it in the legal field. This phase consists of implementing a heterogeneous corpus including legal data, notably articles from the Code of Civil Procedure and the Code of Criminal Procedure, rulings from the Moroccan Court of Cassation and other judicial cases. For training, the Masked Language Modeling (MLM) technique was used to mask tokens in the input data. To do this, 15% of the tokens were selected: 80% of them were replaced by the entity [mask], 10% underwent random filling, while 10% remained constant.

TABLE I. EMPIRICAL STATISTICS OF THE DATASET

Type of case	Count	Mean	Std	Median	Min	Max
Payment cases	118	95	63.1	75,5	28	280
Traffic accident cases	97	213	129	173	62	532
Rental cases	76	158	81.3	135	55	344
Miscellaneous cases	38	388	195.4	316	132	828

The training was conducted with a batch size of 4 combined with a maximum length of 256 tokens per sequence. Each batch represents a set of data processed by the model so that it can adjust its parameters. The number of epochs was set to 5, where at each epoch the model performs a complete cycle on the entire dataset. The model was trained using the AdamW optimizer with a learning rate set to 3e-5, warmup_steps set to 100 and weight decay set to 0.01.

C. Experimentation

For the issue of estimating the duration of a court case, two approaches were adopted: the first was structured on a basis combining automatic classification and statistical correspondence. This method is based primarily on the factual descriptions of cases for automatic classification. In principle, Arabic_LegalBERT served as the basis for producing representations that capture the semantic and contextual information of the facts. The representations obtained are implemented in a downstream module to assign each case to the appropriate type. This classification was carried out using a neural network employing RELU functions. Next, to estimate the duration of the case, the present approach establishes a synergistic link between each type of case, resulting from the classification step, and its duration, by exploiting empirical statistics calculated from the training database. The second approach modeled the problem using a similarity approach combined with empirical statistics. In this approach, textual facts are also transformed into embedding representations through Arabic_LegalBERT. The representations obtained are then compared with each other using the cosine similarity method to identify the most similar cases. For the task of estimating the duration of a legal case, the first strategy consists of calculating the weighted average of the durations of the closest cases detected. The objective of this strategy is to give priority to the closest cases by assigning them the highest weighting. The second strategy uses the average duration of the majority case type among the closest cases detected.

D. Evaluation

In this study, two categories of measures were adopted:

1) Performance indicators for the classification approach: The performance of the classification approach was assessed using accuracy, F1-score, recall and precision, calculated with both weighted averages and macro averages.

a) The macro average: to evaluate the model's behavior across all classes in a balanced manner.

b) The weighted average: to evaluate the overall performance of the models, considering the actual distribution of the data.

2) Performance indicators for the similarity approach: The performance of the similarity approach was evaluated using the following metrics.

a) Mean Absolute Error (MAE): The mean of the absolute differences between predicted values and true values.

$$MAE = \frac{\sum_i^n \left| y_i - y_{i_{préd}} \right|}{n} \quad (1)$$

b) Root Mean Squared Error (RMSE): The square root of the mean of the squared errors, commonly used to emphasize larger errors.

$$RMSE = \sqrt{\frac{\sum_i^n \left(y_i - y_{i_{préd}} \right)^2}{n}} \quad (2)$$

c) Coefficient of determination R²: A statistical measure that quantifies the proportion of variance in the target variable explained by the predictions.

$$R^2 = 1 - \frac{\sum_i^n \left(y_i - y_{i_{pred}} \right)^2}{\sum_i^n (y_i - y_{mean})^2} \quad (3)$$

y_i: Observed value.

$y_{i_{préd}}$: Predicted value.

y_{mean}: Mean of observed values.

V. RESULTS

After calculating the performance metrics of the two fine-tuned models, based on Arabic_LegalBERT and AraBERT, and respectively, as part of the first classification approach, a comparison between the two models was made. The results show that both models perform well across all evaluation metrics, with a slight improvement for the specialized model. The 98% accuracy recorded for Arabic_LegalBERT is 2% higher than the 96% achieved by AraBERT, illustrating the advantage of pre-training on the legal corpus (see Table II). This finding is consistent with previous work that has demonstrated the effectiveness of models pre-trained on specific domains for a set of natural language understanding tasks [3]. Generally, minority classes are well handled, as evidenced by performances exceeding 90% for all metrics. However, there is a slight improvement in weighted measures compared to macro measures, which still indicates an influence of the majority classes.

The cosine similarity-based method was implemented using only Arabic_LegalBERT. Table III summarizes the performance indicated by the MAE, RMSE, and R^2 measures for the two approaches: similarity-weighted average and majority-type average. The mean absolute error indicates that, on average, the predictions deviate from the actual values by approximately 57 days for both approaches. The RMSE indicates an overall dispersion of errors of approximately 79.50 days for the similarity-weighted average approach and 78 days for the majority type-based approach. The coefficient of determination indicates that the model captures 65.21% and 66.56% of the variability in the observed data, respectively. The results of the two approaches are generally close and acceptable, but not very efficient. The model still has considerable potential for optimization, which could be exploited by enriching and increasing the size of the data.

When examining each type of case individually (Table IV), a comparison between the two approaches shows that, overall, the similarity-weighted average approach is superior to varying degrees. The disparity between the two approaches in terms of MAE is evident for rent cases and miscellaneous cases, with a difference of 3 days and 4 days between the two approaches, respectively. Whereas for payment cases, the difference between the two approaches is only 1 day. According to the distribution of our corpus presented in Table I, payment cases represent the majority class with 118 cases, while rent cases and miscellaneous cases constitute the minority classes with 76 and 38 cases, respectively. This configuration could explain these results, as the difference between the two approaches in terms of MAE increases as the number of cases decreases. The absence of similar cases in the minority classes thus seems to reveal the impact of weighting. Conversely, we note that for cases related to traffic accidents, the approach based on the majority type average stands out, with an MAE of 69.5 days compared to 74 days for the similarity-weighted average. This could be explained by the fact that a high degree of textual similarity assigns high weights to cases with very dispersed durations.

VI. DISCUSSION

The exploration of language models applied to the legal field has proven effective. The estimation of judicial case durations is based on representations produced by AraBERT and Arabic_LegalBERT. The two models were fine-tuned for the classification task, and the case duration was then calculated based on historical data corresponding to the predicted case type. The results show that both models perform well, with a slight advantage for the specialized model. This demonstrates the robustness of the original Arabic model [4], as well as the value of additional pre-training on the legal corpus. This finding is consistent with study [3], in which the specialized AraLegal-BERT model, developed for the Arabic legal domain, demonstrated superior performance compared to other language models. Furthermore, although the class imbalance was not very pronounced due to the limited size of the data, it nevertheless had an impact on the performance of both models. The distinction between the specialized model and the basic model could have been more pronounced with a larger pre-training corpus and a greater number of more diverse cases.

TABLE II. PERFORMANCE OF CLASSIFICATION APPROACH

Model	Average (Macro/ weighted)	Accuracy	Precision	Recall	F1-Score
AraBERT	macro avg	96	96.76	90	92.07
	weighted avg	96	96.26	96	95.57
Arabic_LegalBERT	macro avg	98	98.68	95	96.55
	weighted avg	98	98.11	98	97.92

TABLE III. PERFORMANCE OF SIMILARITY APPROACH

Method	MAE	RMSE	R^2
Similarity-Weighted Average	56.58	79.50	65.21
Average by Majority Type	57	77.94	66.56

TABLE IV. PERFORMANCE OF THE SIMILARITY-BASED APPROACH BY CASE TYPE

Type of case	Method	MAE	RMSE
Payment cases	Weighted Average by Similarity	23.61	41.44
	Average_Majority Type	24.85	40.46
Traffic accident cases	Weighted Average by Similarity	74.12	94.16
	Average_ Majority Type	69.50	85.88
Rental cases	Weighted Average by Similarity	59.50	69.09
	Average_ Majority Type	62.77	72.41
Miscellaneous cases	Weighted Average by Similarity	108.35	130.29
	Average_Majority Type	112.92	133.61

In the similarity approach, the results are generally acceptable with an MAE of 57 days. However, this average error remains significant given that the average duration of the cases studied in this work is 178 days. When each type is considered independently, payment cases, which are the most numerous (118 cases), have a MAE of 24 days, the lowest of

all case types (see Table IV). Traffic accident cases (97 cases) and rent cases (76 cases) have MAE of 74 and 59.50 days, respectively. Although the number of rental disputes is lower than the number of traffic accident cases, their lower MAE can be attributed to the reduced variability of cases in this category, as evidenced by the standard deviation 81 compared to 129 for traffic accident cases. On the other hand, miscellaneous cases, which are in the minority with only 38 cases and are very heterogeneous, with a standard deviation of 195.4, have the highest MAE of 108 days.

These shortcomings are mainly due to the limited number of cases and the presence of similar cases that were processed over different durations. Because of these limitations, the model is unable to effectively differentiate between normal cases and outliers in terms of durations or ensure fine-grained consistency in its predictions.

VII. Conclusion

In summary, this work focused on a specific task in the legal field involving the estimation of judicial case durations. This study leveraged language models, particularly transfer learning, to specialize a language model, Arabic_LegalBERT, on an Arabic legal corpus. The specialized model was used to generate contextual representations adapted to this framework. To predict legal case duration, two hybrid systems were adopted: the first combines representations produced by the pre-trained Arabic_LealBERT model and classification with empirical statistics, and the second also combines representations produced by the pre-trained model and the cosine similarity technique with statistical correspondence. In the classification-based approach, the estimated duration was based on historical durations for the type of case to which the case was assigned. The results highlight the performance of both models, with an advantage for the pre-trained model. In the similarity approach, only the specialized model was applied. The duration of a judicial case was estimated based on the closest identified cases. At this level, two variants were adopted: weighted average and majority type average. The results are promising, but the limited data prevents the model from reducing errors and predicting durations with much greater accuracy. This study also highlighted the impact of dataset imbalance on performance metrics. Considering the results obtained and the limitations encountered, future projects should focus on enhancing the dataset by incorporating a larger and more diverse set of cases, as well as exploring other language models, such as LLaMA, and comparing them with BERT-based models, particularly in their specialized versions. Future perspectives also include the development of a specialized model to predict the reasonable duration of a legal case by identifying factors.

Acknowledgment

This work was supported by the Ministry of Higher Education, Scientific Research and Innovation, the Digital Development Agency (DDA) and the CNRST of Morocco (Alkhawarizmi/2020/25).

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin,., « Attention is all you need », Advances in neural information processing systems, vol. 30, 2017.

[2] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, and I. Androutsopoulos, « LEGAL-BERT: The Muppets straight out of Law School », 6 octobre 2020, arXiv: arXiv:2010.02559. doi: 10.48550/arXiv.2010.02559.

[3] M. Al-qurishi, S. Alqaseemi, and R. Souissi, « AraLegal-BERT: A pretrained language model for Arabic Legal text », in Proceedings of the Natural Legal Language Processing Workshop 2022, N. Aletras, I. Chalkidis, L. Barrett, C. Goanță, and D. Preoțiuc-Pietro, Éd., Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics, déc. 2022, p. 338-344. doi: 10.18653/v1/2022.nllp-1.31.

[4] W. Antoun, F. Baly and H. Hajj, « AraBERT: Transformer-based Model for Arabic Language Understanding », in Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, H. Al-Khalifa, W. Magdy, K. Darwish, T. Elsayed, and H. Mubarak, Éd., Marseille, France: European Language Resource Association, mai 2020, p. 9-15.

[5] N. Aletras, D. Tsarapatsanis, D. Preoţiuc-Pietro, and V. Lampos, « Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective », PeerJ Comput. Sci., vol. 2, p. e93, oct. 2016, doi: 10.7717/peerj-cs.93.

[6] M. Medvedeva, M. Vols, and M. Wieling, « Using machine learning to predict decisions of the European Court of Human Rights », Artif Intell Law, vol. 28, no 2, p. 237-266, juin 2020, doi: 10.1007/s10506-019-09255-y.

[7] O.-M. Sulea, M. Zampieri, M. Vela, and J. van Genabith, « Predicting the Law Area and Decisions of French Supreme Court Cases », arXiv:1708.01681 [cs], août 2017.

[8] H. Zhong, Z. Guo, C. Tu, C. Xiao, Z. Liu, and M. Sun, « Legal Judgment Prediction via Topological Learning », in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium: Association for Computational Linguistics, 2018, p. 3540-3549. doi: 10.18653/v1/D18-1390.

[9] W. Yang, W. Jia, Xi. Zhou, and Y. Luo, « Legal Judgment Prediction via Multi-Perspective Bi-Feedback Network », Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, p. 4085-4091, août 2019, doi: 10.24963/ijcai.2019/567.

[10] S. Long, C. Tu, Z. Liu, and M. Sun, « Automatic Judgment Prediction via Legal Reading Comprehension », in Chinese Computational Linguistics, M. Sun, X. Huang, H. Ji, Z. Liu, and Y. Liu, Éd., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, p. 558-572. doi: 10.1007/978-3-030-32381-3_45.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, « BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding », in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), J. Burstein, C. Doran, and T. Solorio, Éd., Minneapolis, Minnesota: Association for Computational Linguistics, juin 2019, p. 4171-4186. doi: 10.18653/v1/N19-1423.

[12] I. Beltagy, K. Lo, and A. Cohan, « SciBERT: A Pretrained Language Model for Scientific Text », 10 septembre 2019, arXiv: arXiv:1903.10676. doi: 10.48550/arXiv.1903.10676.

[13] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, « BioBERT: a pre-trained biomedical language representation model for biomedical text mining », Bioinformatics, vol. 36, no 4, p. 1234-1240, 2020.

[14] Y. Yang, M. C. S. UY, and A. Huang, « FinBERT: A Pretrained Language Model for Financial Communications », 9 juillet 2020, arXiv: arXiv:2006.08097. doi: 10.48550/arXiv.2006.08097.

[15] M. Polignano, P. Basile, M. De Gemmis, G. Semeraro, and V. Basile, « AlBERTo: Italian BERT language understanding model for NLP challenging tasks based on tweets », GERMANIA, 2019.

[16] L. Martin, T. Muller, P. Suárez, Y. Dupont, J. Romary, E. de la Clergerie, M. Seddah, and B. Sagot,« CamemBERT: a Tasty French

Language Model », in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, p. 7203-7219. doi: 10.18653/v1/2020.acl-main.645.

[17] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, and J. Pérez, « Spanish Pre-Trained BERT Model and Evaluation Data », in PML4DC at ICLR 2020, 2020.

[18] M. Abdul-Mageed, A. Elmadany, and E. M. B. Nagoudi, « ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic », 7 juin 2021, arXiv: arXiv:2101.01785. doi: 10.48550/arXiv.2101.01785.

[19] I. Chalkidis, I. Androutsopoulos, and N. Aletras, « Neural Legal Judgment Prediction in English », 5 juin 2019, arXiv: arXiv:1906.02059. doi: 10.48550/arXiv.1906.02059.

[20] P. Bambroo and A. Awasthi, « LegalDB: Long DistilBERT for Legal Document Classification », in 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), févr. 2021, p. 1-4. doi: 10.1109/ICAECT49130.2021.9392558.

[21] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, « DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter », arXiv preprint arXiv:1910.01108, 2019.

[22] S. Douka, H. Abdine, M. Vazirgiannis, R. El Hamdani, and D. Restrepo Amariles, "JuriBERT: A Masked-Language Model Adaptation for French Legal Text," in Proceedings of the Natural Legal Language Processing Workshop 2021, Punta Cana, Dominican Republic, Nov. 2021, pp. 95–1011.

[23] D. Betancur Sánchez, N. Aldama García, Á. Barbero Jiménez, M. Guerrero Nieto, P. Marsà Morales, N. Serrano Salas, C. García Hernán, P. Haya Coll, E. Montiel Ponsoda, and P. Calleja Ibáñez, « MEL: Legal Spanish Language Model », 27 janvier 2025, arXiv: arXiv:2501.16011. doi: 10.48550/arXiv.2501.16011.

[24] D. Licari and G. Comandè, « ITALIAN-LEGAL-BERT: A Pre-trained Transformer Language Model for Italian Law. », Ekaw (companion), vol. 3256, 2022.

[25] A. Ammar, A. Koubaa, B. Benjdira, O. Nacar, and S. Sibaee, « Prediction of arabic legal rulings using large language models », Electronics, vol. 13, no 4, p. 764, 2024.