Available online @ https://jjem.jnnce.ac.in
https: www.doi.org/10.37314/JJEM.SP0249
Indexed in International Scientific Indiexing (ISI)
Impact factor: 1.395 for 2021-22
Published on: 08 December 2024

AI-Based Contract & Legal Document Generator using Machine Learning

Ashwini S I[1] & Santhosh S G[2]

1 Student, Dept. of MCA, JNN College of Engineering, Shimoga
2 Associate Professor, Dept. of MCA, JNN College of Engineering, Shimoga
ashwiniindreshs@gmail.com, santhoshsgrao@jnnce.ac.in

Abstract

The legal document generation can be defined as an autonomic process of drafting legal documents wherein for any person or any legal user whether professional or lawyer, it can provide the right references and decisions. With the help of AI and machine learning, it is possible to generate legal documents in this paper. Examples of legal records include rental agreements, partnership agreements, loan agreements among others. For example, employing principles such as the Convolution Neural Network model and the Long Short-Term Memory model is wonderfully useful to lawyers as it creates legal writing. In a way, this has made things easier because it has reduced the amount of time that lawyers need to spend preparing their cases for trial. Besides that, it enhances the quality of case development since the best references and appropriate judgments are obtained. It also assists judges in making right decisions most especially in courtrooms and frequently during trials. AI and ML have indeed evolved significantly, especially in the recent past and they are increasingly being deployed in many areas including the legal field. this article presents the information about how the legal systems incorporating the artificial intelligence and machine learning technologies are deployed in the development of legal forms. In this paper, we demonstrate this by real case analysis and test results of AI as it is always excellent in coming up with a several legal documents such as contracts and wills among others

Keywords: *Text analysis, Convolution neural network model [CNN], Long Short-Term Memory Model [LSTM].*

1. Introduction

India's legal system is developing extremely slowly. The Supreme Court of India has an astounding 71,411 pending cases that need to be resolved. Of these, 15,076 are criminal cases and 56,365 are civil cases. The High courts in India take an average of only 300 to 500 seconds for one case hearing, puts a lot of pressure on lawyers and Legal Professionals as they gear up for their case studies and trials. Even a tiny mistake could be the deciding factor between winning or losing the case. So sometimes, justice might get overlooked due to human error. The best thing right now is to prolong the duration of the case to gather solid evidence and avoid making incorrect judgements. Artificial intelligence and machine learning are here to help lawyers by reducing human error and finish while additionally giving better portrayal to clients. There are no standard computerization techniques executed in the present Indian overall set of laws that assists attorneys and legitimate inefficiencies through case studies. This translates to quicker case or trial resolution, better judgment presentation through infographics, and less strain on attorneys and advocates. you can save a lot of money on legal document preparation by reducing the amount of physical labour needed for the majority of legal document preparation. This study delves deeply into the application of machine learning and artificial intelligence to legal aid. Lawyers and advocates will use the AI-based prediction model as a guide to determine the necessary level of expertise for each case. The inspiration for this exploration is to distinguish ideal procedures that could support the improvement of an artificial intelligence-controlled Partner for lawful use, which could accelerate the course of case
experts with planning for the portrayal of their client. Because of the great number of cases that are forthcoming in Indian legal courts, this issue enhances moreover and could prompt under

representation because of human limited capacities, expanding time in goal of cases and making the overall set of laws far off for most of the residents. As a country with a masses of over 1.3 billion masses, it is essential to improve the assets accessible in the general arrangement of regulations so every individual can rehearse the significant freedoms guaranteed by the Indian Constitution.

2. Objectives

Efficiency: The drafting method should be automated to save time and lessen the workload for the legal fraternity members, thereby preventing such traumatic experiences.

It is vital to maintain consistency in the form and content of the legal files to reduce the possibility of individual error.

Accuracy: Boost the readability of lawful terminology & ensure that it complies with all applicable laws and parameters.

Accessibility: Making sure authorized services are easily available to everybody, even those who cannot afford to hire an expensive lawyer, is one role that can be played. Legal services can be easily accessed by individuals and organizations, particularly by those who are unable to pay for legal representation.

3. Literature Review

On the basis of “AI Based Contract & Legal Document Generator using Machine Learning”, various work has been completed from different Authors like Rithik Raj Pandey et al. in [1], who detailed the methodology involves data collection & machine leaning techniques are used to train Custom Trained GPT using natural language processing and pattern recognition. the system utilizes optical Character Recognition technology which will examine legal documents PDFs to facilitate relevant information extraction by custom trained GPT. V. Vaissnave et al. in [2], this paper contains data preprocessing step objective of the paper is to create a chatbot and evaluate a chatbot using large language models [LLMs] like Chat-GPT, which can provide precise legal validation to members of Palestinian organizations. This includes responding to questions associated with cooperative laws, bylaws, and guidelines professionally and effectively. Anupama C comprises token embedding & word embedding, model development includes Bi-LSTM (Bidirectional Long Short-Term Memory), LSTM-CRF (Conditional Random Fields), and CNN (Convolutional Neural Networks), for comparison they used SVM classifiers, and numerous RNN methods comprising GRU were evaluated. Dipankar Chakrabarti et al. in [3], This paper presents the "risk-o-meter," a system for examining risks in appropriate archives utilizing machine learning (ML) and familiar dialect making (NLP). Content Representation, Content Classification, Preparing Information, Persistent Learning, and Passage Classification into Risk Categories are a few of the steps in the preparation. Back Vector Machines (SVM) and Naïve Bayes classification calculations are utilized for chance expectation. The models experience assessment on the idea of precision, accuracy, review, and F1-score. PV-DM with negative assessing, the show that performs the best, is chosen for sending. 91% exactness was achieved in the chance category utilizing the biggest preparing dataset. Suma R. et al. in [4], this paper provides a legal assistance which is a computerized approach provides precise references, predictions and decisions to lawyers or any users. They used the models like CNN, which uses K-fold cross validation technique for the evaluation of the model. And it also uses KNN for the prediction of crime. Precia Jacey et al. in [5], This study believes to examine the Execution of artificial intelligence in the field of regulation, remembering how the improvement of artificial intelligence guidelines for China, the US, and Indonesia. Here the author examines the use of artificial intelligence in the field of legal services and focuses on the type of AI used in fields. Rabee Al-Qasem et al. in [6], This paper gives the growth and estimation of a chatbot planned to assist with legal inquiries related to Palestinian laws. The fundamental

Emmanuel et al. in [7], this paper aims to explore and inspect whether ai holds the prospect to alter the legal industry possibility of better performance of its accessibility accuracy and effectiveness in the resolution of legal procedures via ai contracts action framework including interviews concept definitions and case descriptions questions and cases and cross-

sectional studies respectively. Asst Prof Asadullah Shaikh et al. in [8], this paper demonstrates that the author mainly focuses on the application of Artificial intelligence to create & manage legal documents. This paper demonstrates an AI Tool that helps laws, legal professionals or any users in creating legal documents. they used Natural language processing(NLP) including techniques such as tokenization, named entity recognition(NER) & POS widely. Alfredo Montelongo et al. in [9] this study examines the use of deep learning in the legal area tools such as Vos viewer and bib excel are utilized to analyze obtained data models like RNN and LSTM i.e. recurrent neural network long short-term memory is employed to identify pertinent portions in the text is utilized for legal text analysis and record classification data organized logically and sequentially. Parth Shah et al. in [10] this paper concentrates on building and optimizing the process which uses the methods of ml accompanied by heuristics to learn legal provisions from contracts the work might have used various machine learning models like natural language processing methods random forest conditional random fields CRFs support vector machines SVM etc to sum up the paper provides a comprehensive approach for the identification of legal clauses from legal textual content this is accomplished through the synergy of the enhanced LSTM with highly-developed machine learning procedures and other heuristics to achieve a viable and effective system.

4. Proposed Methodology:

proposed system an AI and machine learning-based legal record generator will revolutionize the legal business and make it more efficient resulting in decreased prices and fewer mistakes by automating the creation of legal forms. some of the technologies used are machine learning frameworks such as TensorFlow and PyTorch, NLP libraries which include SPACy and NLTK. this kind of technology brings aids such as minimized human intrusion, Cost-effective document production and accessing of legal facilities.

Datasets: Four thousand eight hundred and twenty five total legal articles are used as a datasets for the analysis & generation of legal documents.

Steps or workflow involved in creating legal documentation involves Data Collection, Data Preprocessing, Feature Engineering, Model Selection and Training, Validation & testing, User Interface and Integration, Deployment, Monitoring and Maintenance.

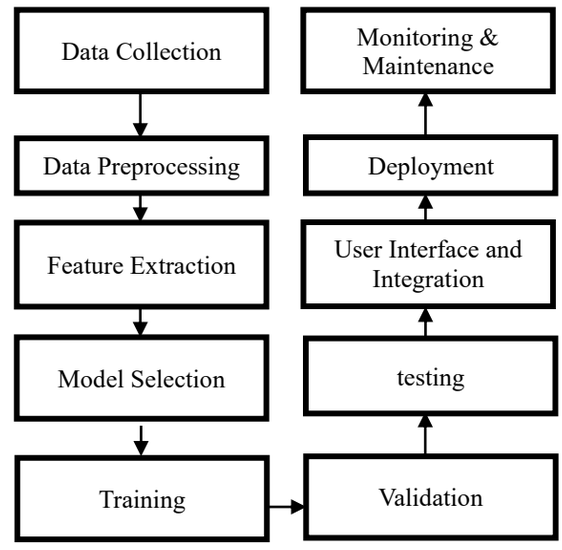


Fig 4.1: Workflow of Proposed methodology

Data Collection: Compile a thorough dataset of court cases that includes the facts, arguments made in support of the cases, and the rulings that accompanied them.

Data Preprocessing: Organize and prepare the legal texts by eliminating superfluous details, harmonizing formatting, and attending to any data that is absent.

Feature extraction: Extract pertinent features from the legal texts so that the model can be trained with them. This could entail gathering pertinent data such as legal citations and keywords.

Model Selection: Select a machine learning model that is suitable for the given task.

Training: Use the labelled dataset to train the chosen model. In order to reduce prediction errors, this entails feeding the model with features taken out of the legal texts and modifying its parameters.

validation: To make sure the model generalizes well to fresh, untested data, assess its performance on a different validation dataset.

Testing: To replicate the model's real-world applicability, evaluate its performance on a

different test dataset.

Model deployment: After the model runs well, move it into a real-world setting. This can entail incorporating it into a specialized prediction tool or a platform for legal research.

Monitoring and Maintenance:

Consistently assess how well the system performs in practical situations. Periodically add new data to the model so that it can adjust to changing legal trends.

Machine Learning and AI techniques that are used in generating legal documents are:

- Text Analysis
- Convolution neural network model
- Long Short-Term Memory Model

4.2. Text Analysis:

text analysis involves reading and digesting the material followed by extracting the legal text or the keywords from the record the method that is employed in text analysis is known as NLP or natural language processing.

Various NLP procedures we use for data preprocessing include tokenization, stop word elimination, stemming and parts of speech tagging.

4.2.1. stop words:

These are the terms which are frequently used in any text of that language. In data mining and text filtering, stop words are used so as to separate the frequently used terms so that the unusual words can be identified. Some of the examples of the English stop words are the, is, are, if, can, and so on.

4.2.2. stemming:

Stemming is a basic data pre-processing method in which a derived or inflicted term is brought back to its shortest possible form. The word stemming in turn derives from the base word, which is also referred to as the stem word. In data processing and analysis, stemming is quite important because while the meaning of a word remains the same across contexts, a text file may have many different forms of stem word like the inclusion of suffixes or prefixes [19]. It would be unreasonable to treat every word with the same stem as a separate word as it would also be a waste of time and computational resources.

4.2.3. tokenization:

During the pre-processing step, the words, characters, terms etc. of the text is segregated in small parts each of which is called a token and each piece or block is also called token. One of these stages is the tokenization stage, which is counted to be the most basic level NLP. It matters because people read texts and are able to differentiate between different word part types. It will also further the text illustration so that the text's meaning can be analyzed. Other needed tokens come with other Natural Language uses, such as sentiments and removal of unnecessary words called stop words.

4.2.4 parts of speech tagging:

By using NLTK (As the name suggests NLTK is a package in python used for natural language processing) it is easy to tokenize the text on the basis of the part of speech that they really belong to. This type of tokenizing can greatly assist the machine not only in considering word by word text scan but also the connotation of the message that is being passed.

4.3. Phrases based LSTM-CNN model Architecture:

Input & Output:

The LSTM-CNN model is a sequence-to-sequence model that uses phrases as input and output sequences. Phrases can be divided into three major types: subject phrases, relational phrases, and object phrases.

Convolutional phrase encoder:

The model uses a convolutional neural network model for representing phrases, which can be trained efficiently and have been successfully used for sentence-level classification tasks.

LSTM:

The recurrent document encoder is a model that uses LSTM instead of gated recurrent units (GRU) due to its easier to tune parameters and stronger theoretical guarantee.

Decoder:

The decoder part of the model divides into two modes: generate mode and copy mode. In generate mode, the decoder predicts the next phrase based on all annotations obtained in encoding and all previously predicted phrases. In copy mode, the decoder guesses the next sentences based on the absolute value of the maximum value of the maximum value. The model uses a log loss function for loss function.

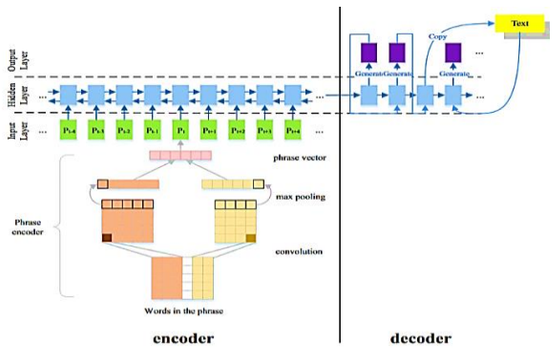


Fig 4.3.1: Architecture of Phrases based CNN-LSTM model

5. Result

```
14/14 [==============================] - 12s 397ms/step - loss: 2.6867 - accuracy: 0.7176 - val_loss: 1.4397 - val_accuracy: 0.7907
Epoch 2/10
14/14 [==============================] - 4s 290ms/step - loss: 1.4056 - accuracy: 0.7938 - val_loss: 1.3930 - val_accuracy: 0.8164
Epoch 3/10
14/14 [==============================] - 4s 290ms/step - loss: 1.3054 - accuracy: 0.8187 - val_loss: 1.3308 - val_accuracy: 0.8344
Epoch 4/10
14/14 [==============================] - 4s 290ms/step - loss: 1.2496 - accuracy: 0.8344 - val_loss: 1.3063 - val_accuracy: 0.8343
Epoch 5/10
14/14 [==============================] - 4s 290ms/step - loss: 1.2226 - accuracy: 0.8353 - val_loss: 1.3141 - val_accuracy: 0.8347
Epoch 6/10
14/14 [==============================] - 4s 290ms/step - loss: 1.2086 - accuracy: 0.8353 - val_loss: 1.3109 - val_accuracy: 0.8347
Epoch 7/10
14/14 [==============================] - 4s 289ms/step - loss: 1.1918 - accuracy: 0.8354 - val_loss: 1.3089 - val_accuracy: 0.8358
Epoch 8/10
14/14 [==============================] - 4s 290ms/step - loss: 1.1685 - accuracy: 0.8354 - val_loss: 1.3044 - val_accuracy: 0.8359
Epoch 9/10
14/14 [==============================] - 4s 290ms/step - loss: 1.1471 - accuracy: 0.8359 - val_loss: 1.3032 - val_accuracy: 0.8362
Epoch 10/10
14/14 [==============================] - 4s 289ms/step - loss: 1.1203 - accuracy: 0.8368 - val_loss: 1.3102 - val_accuracy: 0.8363
```

Fig 5.1: Training accuracy table

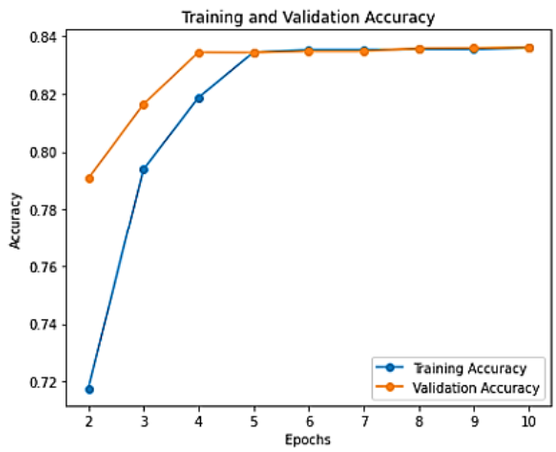


Fig 5.2: Training accuracy graph

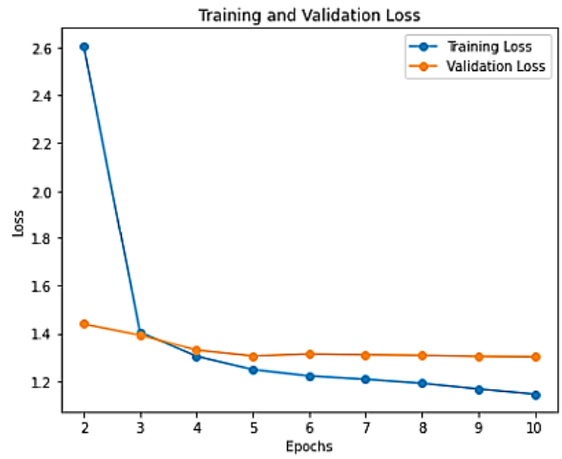


Fig 5.3: Training Loss graph

Fig 5.1, 5.2 & 5.3**:** The execution of the model covered ten epochs, and several indicators, including loss, accuracy, val_loss, and val_accuracy were logged during the process. The model I created displayed a low training loss and a high level of accuracy, meaning that it was able to learn properly from training data. By the 7th epoch though, validation accuracy stabilized at a certain level, indicating over-fitting. Future work should carry out the work on trying to fine tune the model on how to avoid getting over-fitted.

Fig 5.4: Login page

For Generating legal documents, the web page is created, in the login page is designed.
In the login page, user or any legal professional has to enter his details then click on login. If he is a new user he has to sign up using credentials.

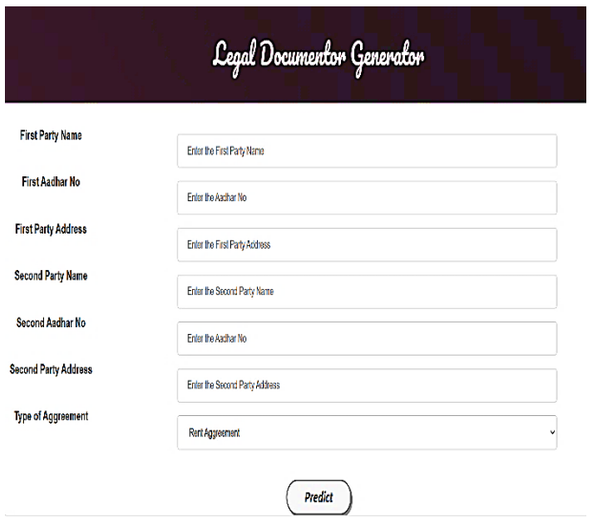
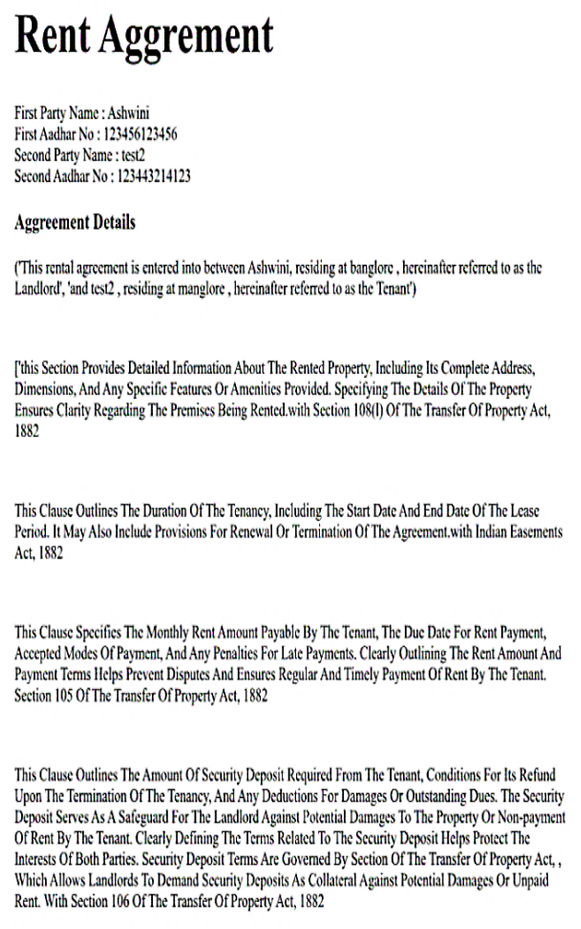


Fig 5.5: Legal document form

After user login, legal document form will comes up where user has to enter First party's name, Aadhar number & address. same goes to second party also. Then click on predict button.



Rent Aggrement

First Party Name : Ashwini
First Aadhar No : 123456123456
Second Party Name : test2
Second Aadhar No : 123443214123

Aggreement Details

('This rental agreement is entered into between Ashwini, residing at banglore , hereinafter referred to as the Landlord', 'and test2 , residing at manglore , hereinafter referred to as the Tenant')

['this Section Provides Detailed Information About The Rented Property, Including Its Complete Address, Dimensions, And Any Specific Features Or Amenities Provided. Specifying The Details Of The Property Ensures Clarity Regarding The Premises Being Rented.with Section 108(I) Of The Transfer Of Property Act, 1882

This Clause Outlines The Duration Of The Tenancy, Including The Start Date And End Date Of The Lease Period. It May Also Include Provisions For Renewal Or Termination Of The Agreement.with Indian Easements Act, 1882

This Clause Specifies The Monthly Rent Amount Payable By The Tenant, The Due Date For Rent Payment, Accepted Modes Of Payment, And Any Penalties For Late Payments. Clearly Outlining The Rent Amount And Payment Terms Helps Prevent Disputes And Ensures Regular And Timely Payment Of Rent By The Tenant. Section 105 Of The Transfer Of Property Act, 1882

This Clause Outlines The Amount Of Security Deposit Required From The Tenant, Conditions For Its Refund Upon The Termination Of The Tenancy, And Any Deductions For Damages Or Outstanding Dues. The Security Deposit Serves As A Safeguard For The Landlord Against Potential Damages To The Property Or Non-payment Of Rent By The Tenant. Clearly Defining The Terms Related To The Security Deposit Helps Protect The Interests Of Both Parties. Security Deposit Terms Are Governed By Section Of The Transfer Of Property Act, , Which Allows Landlords To Demand Security Deposits As Collateral Against Potential Damages Or Unpaid Rent. With Section 106 Of The Transfer Of Property Act, 1882

This Clause Details The Procedure For Terminating The Tenancy, Including The Required Notice Period For Both Parties And Conditions Under Which The Agreement May Be Terminated. Indian Easements Act, 1882

First Party

Sign & Seal

Second Party

Sign & Seal

Fig 5.6: Rent Agreement Template

After clicking on the predict button, the above rent agreement will be generated which consists of the first & second party names, addresses and other details at the beginning & agreement details which consist of all the clauses & conditions, followed by both parties' names, sign & seal.

6. Conclusion

The use of artificial intelligence and machine learning in the preparation of legal documents is thus considered a major improvement in productivity and quality in the legal area. this technology enables users or any legal professionals to manage the drafting process effectively minimize mistakes as well as improves adherence to legal requirements, this technology is likely to be more innovative in forthcoming and this will help to transform the practice of law because lawyers will only be involved in the provision of added value services such as consulting and client business relations.

7. Reference

1. Rithik raj pandey, legalseva - AI-powered legal documentation assistant, International Research Journal of Modernization in Engineering Technology and Science, e-ISSN: 2582-5208, Volume:06 Issue:03, March-2024

2. V. Vaissnave, An Artificial Intelligence based Analysis in Legal domain, International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-9 Issue-2S2, December 2019

3. Dipankar Chakrabarti, Use of Artificial Intelligence to Analyse Risk in Legal Documents for a Better Decision Support, Proceedings of

TENCON 2018 - 2018 IEEE Region 10 Conference (Jeju, Korea, 28-31 October 2018)

4. Suma R, Review and Approaches to Develop Legal Assistance for Lawyers and Legal Professionals using Artificial Intelligence and Machine Learning, International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395-0056 Volume: 09 Issue: 12, Dec 2022

5. Precia Jacey, Artificial Intelligence: Implementation in Legal Services (Comparative Study on China, United States and Indonesia), Proceedings of the 3rd Asia Pacific International Conference on Industrial Engineering and Operations Management, Johor Bahru, Malaysia, September 13-15, 2022

6. Rabee Al-Qasem, Towards the exploitation of LLM-based chatbot for providing legal support to palestinian cooperatives, arxiv:2306.05827v1 [cs.CL] 9 Jun 2023

7. Anupama C Emmanuel, AI Contract: Potentiality of AI Contract to Revolutionize the Legal Field, International Journal of Research Publication and Reviews, ISSN 2582-7421, Vol 4, no 9, pp 2090-2093, September 2023

8. Asst Prof Asadullah Shaikh, AI-Powered Legal Documentation Assistant, Journal of Emerging Technologies and Innovative Research (JETIR), ISSN-2349-5162, Volume 11, Issue 4, April 2024

9. Alfredo Montelongo, Tasks performed in the legal domain through Deep Learning: A bibliometric review (1987-2020), 2020 International Conference on Data Mining Workshops (ICDMW) | 978-1-7281-9012-9/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICDMW51313.2020.00113

10. Parth Shah, Legal Clause Extraction from Contract Using Machine Learning with Heuristics Improvement, 4th International Conference on Computing Communication and Automation (ICCCA), 978-1-5386-6947-1/18/$31.00 ©2018 IEEE, 2018.