Scope of Artificial Intelligence in Law

Jyoti Dabass 1,* and Bhupender Singh Dabass 2, *

1 Department of Electrical and Electronics Engineering, The Northcap University, Gurugram, Haryana; jyoti16ecd007@ncuindia.edu
2 Department of Law, Institute of Law and Research, Faridabad, Haryana; 93bhupenderdabass@gmail.com
* Correspondence: jyoti16ecd007@ncuindia.edu, 93bhupenderdabass@gmail.com; Tel.: +91 9728202799

Abstract: Over the years, artificial intelligence (AI) is spreading its roots in different areas by utilizing the concept of making the computers learn and handle complex tasks that previously require substantial laborious tasks by human beings. With better accuracy and speed, AI is helping lawyers to streamline work processing. New legal AI software tools like Catalyst, Ross intelligence, and Matlab along with natural language processing provide effective quarrel resolution, better legal clearness, and superior admittance to justice and fresh challenges to conventional law firms providing legal services using leveraged cohort correlate model. This paper discusses current applications of legal AI and suggests deep learning and machine learning techniques that can be applied in future to simplify the cumbersome legal tasks.

Keywords: Legal artificial Intelligence; Machine Learning; Deep Learning; Image Processing; Matlab

1. Introduction

Legal AI can be visualized as the exercise of technologies such as machine learning, natural language processing, speech recognition, legal robotics, planning, natural image understanding, rules-based expert system, neural networks, logic programming, artificial vision, machine learning and neural networks in relation to legal issues. The concept of AI gathered huge importance because of its ability to deal with a large amount of data. Results of AI are more accurate because of inclusive plus apposite along with fast testing making it a hot topic in the field of legal practice. Manual dealing with resources consume time and involves a large amount of money making clients unhappy. Legal AI has shown tremendous results in different applications including Case-based reasoning, document modeling, deontic logic, conceptual retrieval, intelligent tutoring and so on. [1-4].

A subset of AI is Machine learning in which learning of neural networks occurs by analyzing the huge amount of statistics. It is apposite for problems involving vocalizations detection, astuteness and image recognition where mathematical equations and mathematical models fail to produce a final model.

Deep learning is a variety of appliance learning in which multiple neural networks are connected together in order to arrive at nuance answers or to gain deeper context [5-6]. This paper deals with current applications of AI technologies along with challenges faced and future scope of art in the field of law. Section 1 introduces the topic followed by challenges faced by legal AI and proposed solution in section 2. Section 3 involves the current applications of legal AI whereas section 4 discusses how legal AI challenges human legal proficiency ending with the conclusion in section 5.

2. Legal AI: Challenges and Proposed Solution

Artificial intelligence, Machine learning, and deep learning are allied to each other. It is because of the fact that deep learning is a variety of appliance (machine) learning and machine learning is a class of artificial intelligence. Legal AI includes intelligent interfaces to help lawyers in finishing legal tasks, contract analysis to help people in finalizing contracts by analyzing it in depth and Legal data research for analyzing the legal data. To perform these tasks legal AI uses the concept of machine learning which
involves inputting a large amount of data, learning data by machine in training phase and finally providing outcome based on learning.

AI and Machine learning face a large number of challenges while dealing with legal issues. One of the biggest challenges is to retrieve the judicial opinions out of unstructured or noisy legal data which leads to under and over exclusive results. Also, the percentage of imprecision in outcome increases with the complexity of legal questions. The second challenge that these techniques face is to make a distinction between input data and training data. It is required because corrupted or wrong training data can lead to the false outcome. While making a distinction between two data it must be ensured that the data given for training is unbiased which helps in achieving good generalization performance [7].

Overfitting is the primary cause of poor generalization. To understand the concept of overfitting, consider a problem of classification of data position in order to obtain a curvature that defines the rim of two groups using the concept of training data. Let’s assume that some outliers infiltrate the part of an additional cluster and distressed the boundary. In additional expressions, we can say that data is noisy. When the concept of machine learning is applied to this data, the machine learning will assume full data including the noise and finally give improper curve as an outcome. If for a moment you assume that the training data is correct and fitting the model with precision then the model you achieved will have low generalizability. This is called overfitting. To deduce the concept it can be annotated that reducing error in training data leads to overfitting which spoils or degrades the generalizability. To prevent overfitting you can use validation and regularization [8].

Complex models tend to overfit. To confront overfitting at the small cost of performance regularization build a simple model. For example, in case of classification of data, the model could not classify some points but was able to reflect the overall characteristics of model well. Regularization works well when the training data is simple but for data of high dimensions, we need validation.

In case of validation, training data is divided into two groups, one for validation and one for training. Then the model is trained with training data and feat of the model is evaluated via validation set. If the model workings well then training is finished else the modification of model is continued to get the appropriate model.

Like validation, cross-validation also divides the training data into groups but in addition, it changes the dataset. To maintain randomness of validation dataset, it divides the data repeatedly in order to avoid retaining of initially divided sets. This method works well in confronting overfitting [8-9].

3. Legal AI: Current Applications- Insight up front

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation as well as the experimental conclusions that can be drawn.

Table 1. Current applications of Legal AI

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Application</th>
<th>Description of Application</th>
<th>Related software/ firm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Due Diligence [10]</td>
<td>To uncover background information litigators use AI tools to perform diligence. It helps in advising clients on the options available in a legal situation and the required action to overcome the problem.</td>
<td>Kira Systems, Levertor, eBrevia, Ross Intelligence, JP Morgan, Thought River, Law Geex, Judicata, Legal Robot, Casetext’s CARA.</td>
</tr>
</tbody>
</table>
4. Artificial intelligence challenges human legal proficiency

In the evolution stage of law, analog materials including textbooks, casebooks, case reporters, loose-leaf services, law journals, and legislation were a part of every lawyer’s library. Management of these documents was a tedious task as everything was in hard copy. Their storing, preserving and consulting and sorting required a lot of time.

In order to save manual labor in searching texts and time complexity in dealing with analog things, development of CD-ROM, digital materials, and digital libraries attracted lawyers towards digitization by 1990s. Lawyers and clients started exploring Google, digital libraries like Lexis, Bloomberg, Justia (United States) or CanLII (Canada), and government documents akin to judicial regulations, legislation, and judicial opinions.

Latest attraction of lawyers is towards computational technologies comprising predictive tools, and natural language processing that enables users to extort and repossess the relevant text out of noisy data using search engines, speech-to-speech translations and intelligence assistants for user’s reimbursement. In particular, artificial intelligence is anticipated to interrupt and make over the customer understanding, implementation of management of firm infrastructure and legal processes. AI is proficient to do analytical tasks and reproduce most of the work currently done by humans. Critical here is the capacity to incarcerate and manipulate inferred legal knowledge that helps in decision making both by lawyers and their clients. Such systems act as a legal warehouse, building up and learning from their own comprehension to improve the advice offered.

In spite of numerous challenges faced by artificial intelligence, Legal AI is challenging human expertise contributing in legal services by means of legal data research, predictive technology, eDiscovery, intelligent interfaces, triage services and legal bots. Legal data research comprises predictive system where legal artificial intelligence foretell the outcome of a case based on particular theme along with/ or lawsuit leaning based on courtyard conclusion and knowledge scheme of authorized research along practice lines [16].

Predictive Systems are used in prelitigation planning. This AI-driven software inspects a large number of publicly accessible court credentials, cases, and verdict finished by the adjudicator in precedent up to the current day that is connected to the case along, with an innovative variety of valuable community statistics. A predictive system analyzes data which include damage /costs awarded, cases settled by certain companies, triumph/failure charge of certain attorney, winner/collapse of the petition, appealing/ losing point of view, and views/rulings of judges. Its foremost endeavor is to condense the size of labor-intensive study and offer lawyers and clients with actionable impending into prior cases, the deed of lawyers on
analogous matter, and where doable bestow some clue of indemnity that could be awarded by such affair/and/or other feel value information on the root of its collected substantiation in terms of liable triumph of a issue compared to preceding alike matters. Lex Machina provides timing analytics features and artificial intelligence to predict the estimated time of the process of case trial before a specific judge [11].

The simplest technique used for analysis is machine learning which finds out the finest form out of training data which includes legal documents, images, audio etc. In relation to legal AI, we can say we have a legal expert system which is actually an analytical form dependent on know-how along with proficiency of legal proficient thus working as e-lawyer. Machine learning is of two sorts: supervised knowledge which instructs a representation on acknowledged input/output information so that it can envisage upcoming upshot in presence of uncertainty and unsupervised learning which discover concealed prototype or basic formation in input information.

Supervised learning exercise classification and regression procedure to extend prognostic sculpt. Classification techniques foretell unconditional reply by classifying the input data into categories, for example, whether the appeal is success/failure or whether the argument is winning/losing. Typical application includes document/image/video/sound recognition. Regression techniques envisage uninterrupted answer, for example, alteration/fluctuations in claim/damage cost awarded according to the profession. Typical relevance includes the winning/losing rate of lawyers or appeals.

Unsupervised learning finds intrinsic structures or hidden patterns in information. It is useful to portray an inference from dataset comprising of enter information devoid of labeled reactions. For example, you have sumptuous and disarranged legal data including files, e-books of law, magazines and other credentials. Suppose out of all data, you wish to store data related to criminology in one bunch, one related to civil law in another group and so on. Clustering saves manual labor and time by doing it in few seconds. Secondly, we can collect files on the basis of dates or awarded judgments’ to arrange the legal repository.

Selecting the right algorithm for classification and regression in legal issues is partly based on trial and error method. It is because greatly elastic models have a tendency to overfit information by inducing noise. Selecting the exact algorithm needs trading off one gain not in favor of other, counting complexity, accuracy, and speed. MATLAB offers paraphernalia to facilitate you undertake a diversity of appliance learning models and desire the finest [17].

For classification, MATLAB offers different techniques including SVM, Discriminant Analysis, naïve Bayes, nearest neighbor, neural networks. While for regression techniques like linear regression, GLM, SVR, GPR, ensemble methods, decision trees and neural networks work well. Matlab also provides classification and regression learner app which allows us to select the best model. These apps explore the data, select the features, visualize results and thus protects against overfitting using hold out or cross-validation. Using these apps and techniques one can classify our data according to litigations, decisions, judgments and appeals on cases in order to arrange the data in a structured way.

For clustering using machine learning, Matlab first access and load the data followed by preprocessing data and deriving features using preprocessed data. It trains the model using feature derived to find the best model and then integrate the best model to produce results. Matlab offers several techniques for clustering which includes K-means, K-Medoids, Fuzzy C-means, and hierarchal, Gaussian model and neural networks. To eradicate wrong outcomes and classification for refining search software like Brainspace Discovery uses clustering techniques along with sorting techniques. Using these clustering techniques one can make different clusters for sections, taxations, clauses and laws related to specific cases [18].

Knowledge systems trim down the necessitate for lawyers functioning in the position of specialized support lawyers or at least those who are managing somewhat clear-cut research matters by intellectual deep diving into the material available to provide the answers you require. In this, the structure is both learning from the problem an attorney is asking and seeking to deduct the unsurpassed answer from the
information. It does not eliminate the want for scrutiny by the superior lawyer of study. On the other hand, it speed up essential lawful explore done by junior lawyers functioning as a branch of the big team by captivating memo of the assured key, though ‘vanilla’ legal points from credentials that firm desires to append for client remuneration. For this one can use trademark now software to mitigate weak long exertion for registered patents, products and trademarks. ANAQUA Studio also utilizes this knowledge based system to provide literate claims, identify defect along with circular claim references [19]. Legal AI is exploited in great numbers by lawyers in different applications which are discussed below.

4.1. Contract Review

To help people in finalizing the contracts, legal AI first reads and analyzes legal agreement such as leases and commercial contract to extract the meaningful data from them, and/or checking them against current laws/ rules. The simplest soft computing technique that helps in decision making is the fuzzy rule-based system. Fuzzy techniques are based on natural language so works well in dealing with imprecise data. It translates the human solution into fuzzy dependency and common language using fuzzy if-then rules. Some of its applications that law firms found are lease law, eDiscovery, due diligence, sales/ procurement contract review, compliance and risk review, financing/ OTC derivative agreement review and employment contract review.

Natural Language processing is more accurate and faster at reading contracts compared to human lawyers. Natural language processing works well in case of documents while for images in the contract one can go for natural image processing. Using optical character recognition, one can recognize the text from the input image. In Matlab txt= ocr (I) returns an OCR text containing optical character recognition from the input image, I. Here the object contains location, text, a metric and recognized text indicating the confidence of the recognition text. Software like smart shell also uses natural language processing and AI to provide drafting, identifying patent issues, reviewing and formatting documents. Lex Machina’s legal analytics provide information about plaintiff, counsel and their representatives and the sued one to generate the likelihood of winning or losing a case.

Although there are hybrids in AI contract review it can still be said to fall into two main product varieties. First is volume contract review whose objective is to analyze a large number of legal documents to seek out specific legal issues in leases/ contracts in order to give the overall picture to a client of legal status derived from the document. It also helps in finding anomalies i.e. spotting area that requires further legal attention. Second is contract assistance which is useful for non-lawyers who wish to comprehend say 50 pages procurement in pre-signing phase by spotting clauses that other party included. It also helps in re-examine or in adding a legal clause to meet standard internal rules/ practice for the contract. Apart from this, e-discovery uses machine learning and natural language processing to analyze document by making parallel use of AI techniques in contract review. Different companies are using AI tools to simplify their cumbersome legal tasks which are given below.

- Kira Systems uses Legal AI tools to perform due diligence contract review by searching, highlighting and extracting relevant content for analysis.
- Leverton extracts information such as maintenance costs, payable rent and expiration dates from thousands of documents and then organizes them on a spreadsheet which can be used for prediction.
- eBrevia claims that its software using machine learning and natural language processing is able to accurately analyze more than 50 documents in less than one minute time.
- JP Morgan claims that their program named COIN (Contract Intelligence) extracts 150 attributes from 12000 commercial credit agreements in only a few seconds which is equivalent to 36000 hours of legal work by lawyers and their loan workers.
- ThoughtRiver alleges that its software automatically flagging contracts by reading the meaning of clauses provided by AI. AI software validates contracts that are within predefined policies and
suggestions methods for editing and approval by combining text analytics, legal knowledge, machine learning and statistical benchmarks. The software of legal robot uses machine learning and AI to change the legal content into numeric form in order to raise issues on the document.

- ROSS intelligence utilizes natural language search capability which helps in responding questions after searching through billions of documents.
- Casertext’s CARA allows lawyers to foretell an opposing counsel’s arguments by finding opinions previously used by lawyers. It also detects unreliable and negatively treated cases.
- Judicata software read and analyze legal briefs to evaluate their pros and cons in order to assign a score for each brief based on drafting, arguments, and context [20].

4.2. Intelligent interfaces

AI enabled systems to assist lawyers and clients to carry out every day and speedy legal undertaking that entail some experienced guidance to finish. It helps the systems in providing right answers that are customized to user’s needs. The intelligent checklist allows people to have right facts. The software generally utilizes drop-down set of choices and checkboxes to shift the client all the way through sequence of steps so that they can either be specified the accurate information they desire, for example in answer to a unambiguous legal inquiry or be used to seal in the misplaced elements of a set manuscript. They may be apparent facing for clients to employ, or innermost facing, allowing legal representative to issue the proficient system for their own precise necessities and/or facilitate them to finish a officially permitted manuscript. An expert system where inward facing or outward facing, are premeditated in a way to facilitate systems to respond to a assured form of job quarrel, or to assist insert in data to a certain category of legal form [21].

4.3. Deep Learning techniques for image related operations in legal documents

Deep learning is a subdivision of machine learning that educate computers to do what naturally comes to individual i.e. realize from happening. Machine learning algorithms exercise computational methods to find out information unswervingly from information without relying on a prearranged equation as a form. Legal documents include images of witnesses, area and other. Deep learning is especially suited for image acknowledgment which is used for pinpointing such as facial identification, motion recognition. In contracts related to construction, deep learning helps in detecting lane, parking and other relevant information in an image [22].

The simplest way to use deep learning is by pretrained networks whose early layers learn low-level features (edges, blobs, colors) and last layers learn task-specific features. By replacing final layers with new layers one can make system learn features specific to your dataset followed by training network to predict and access network accuracy. Finally, test images are given to trained network to deploy results.

Before applying deep learning techniques, pre-processing is done in initial stage. By pre-processing, we mean dehazing, denoising, degaussing and enhancing the visual quality of image or documents. Using image labeler app for interactive image labeling, one can label region of interest for object detection, pixels for semantic segmentation and scenes for image classification. Its objective is to define ground truth for image collection, train object detection and create automation algorithm for image labeling [23].

Deep learning for computer vision carry out classification, object recognition and transfer learning through convolutional neural networks (CNN or Conv Nets) and semantic segmentation. CNN’s are implemented as a sequence of interconnected layers that are prepared of recurring blocks of convolutional, ReLu (rectified linear units), and pooling layers. The convolutional layers convolve their input with a set of filters that are automatically well-read during network working out. The ReLu layers add non-linearity to networks enabling the network to fairly accurate the non-linear mapping between image pixels and the
semantic content of an image. The pooling layers downsample their inputs and help consolidate local features [24]. Companies using predictive technology are discussed below.

- Everlaw uses the predictive coding feature to classify documents into relevant and irrelevant ones, Together with predictive technology.
- DISCO uses cloud technology to search documents on large data volumes.

A CNN is a well-liked deep learning structural design that repeatedly learns constructive attribute demonstration, directly from picture information. Applications are deep learning can be briefed as follows.

- Bag of features to encode figure features into a squashed depiction apt for image taxonomy and image repossession.
- Template matching applies a small picture, or pattern, to unearth matching regions in a bigger picture.
- Blob analysis exploit segmentation and blob belongings to make out objects of significance.
- The Viola-Jones algorithm utilizes Haar-like features and a cascade of the classifier to categorize things including eyes, faces and noises. You can teach the classifier to distinguish additional objects too [25].

Computer vision system toolbox offered by MATLAB gives pre-trained entity detections and the functionality to instruct a custom detector which uses Viola-Jones algorithm in order to identify people’s faces, noses, eyes, mouth or upper body. The people detectors perceive people in an input picture using the histogram of object-oriented gradients (HOG) features and a trained support vector machine (SVM) classifier. One can modify the cascade object detector utilizing the trainCascadeObjectDetector function [26]. For classification and features matching one can utilize the following concepts.

- To classify images into categories, you can create a histogram of illustration declaration occurrences that symbolizes an image. These histograms, called a bag of visual works, are used to train an image category classifier. Computer vision system toolbox functions also help in the search by image, also known as a content-based image retrieval system which retrieves images from a collection of images that are similar to a query image [27].
- In order to find the unique features out of cluttered scene using objects of non-repetitive texture patterns, one can use point feature techniques. But this technique does not work for homogeneously dyed items, or for stuff having repeating patterns. One can compare the performance of Matlab software with Catalyst and bright flag software which are discussed in brief as under.
- Catalyst software also allows users to lay down patterns such as public safekeeping number on redacted software after removing sensitive and confidential information on documents.
- Bright flag software also provides analytics features to track and categorize data pricing in order to find alternative budgets and fees arrangements.

4.4 Triage Services and legal bots

Using apparatus learning and by processing natural verbal communication, artificial intellect method steer clients in reaching at report outcome in terms of advisory path. On every fresh client dealings at website interfaces, triage structure updates CRM systems to take information, reply queries and inform lawyers about potential client within set service period. It requests emergency contacts from lawyers in order to connect the details provided by way of triage arrangement to firm’s personal study into a brand of case worth shadowing. AOL, Target, and Microsoft use Extern software to carry out e-discovery exertion of law officers at expenditure savings of 95 percent. Legal expert using AI tools for due meticulousness embrace iManage’s RAVN, Seal, Legal sifer and luminance. Stages of triage services include pre-processing, training and testing phase. Success of triage services depend on its accuracy and speed.

Training of algorithm can speed up by tuning hyper-parameters for machine learning algorithms. Bayeopt in MATLAB software selects the optimal machine learning hyperparameters which include learning rate, optimization environment etc. using the technique of Bayesian optimization to train legal documents for further classification and clustering. Parallel optimization helps in searching for the
parameters that give the lowest cross-validation error. After optimizing one get more accurate and speedy outcomes.

On the other hand Legal bots operate bring into play written text which helps in giving advice on matters such as a criminal law to public members or to aid public members through the progression of carrying out a confront to parking payment. Software used for triage services and legal bots are discussed below.

- Smokeball software that auto revises time footage and valid email billings.
- Intraspexion which searches documents with high risks and cluster them according to risk to detect threats of litigation

There is a long list where legal AI is currently used. One cannot predict whether artificial lawyer will replace the need for lawyers or not but can assure it will definitely mitigate manual work to a great extend. Legal AI can appear as bloom especially in countries where judges are less in number and pending cases are adding its number. The main objective of writing this paper is to encourage the young lawyers to go partially artificial by using techniques of Matlab software along with natural language processing [28].

5. Conclusions

Artificial intelligence techniques are gaining importance owing to augmented client demands on ruling firms to be extra proficient and a mounting aversion to recompense for what they consider as progression stage toil. This paper discusses the current trends of legal AI and suggested techniques that can be applied in the time to come. In future, one can come up with the implementation of deep learning and machine learning techniques on legal documents using the Matlab software. Using classification with deep learning we can come up with tax commandment classifier based on in print available Supreme Court, High Court and district court verdict in which courtyard declared twofold resolve for the precise legal issue. Also one can make a classifier by asking a questionnaire based on psychology to define criminals into categories of hard core, general, professional killers and others using their replies to questions.

References

8. Thearling K. An introduction to data mining.
10. Avi-Yonah RS, Mazzoni G. Due Diligence in International Tax Law.
18. Gehrking D, Law C, Maxwell A, inventors; Google LLC, assignee. Categorizing objects, such as documents and/or clusters, with respect to a taxonomy and data structures derived from such categorization. United States patent US 9,971,813. 2018 May 15.