The Impact of Language Technologies in the Legal Domain

Isabel Trancoso, Nuno Mamede, Bruno Martins, H. Sofia Pinto, and Ricardo Ribeiro

Abstract In the current digital era, language technologies are playing an increasingly vital role in the legal domain, assisting users, lawyers, judges, and legal professionals to solve many real-world problems. While open datasets and innovative deep learning methodologies have led to recent breakthroughs in the area, significant efforts are still being made to transfer the theoretical/algorithmic developments, associated with general text and speech processing, into real applications in the legal-domain. This chapter presents a brief survey on language technologies for addressing legal tasks, covering studies and applications related to both text and speech processing (Manuscript submitted in May 2022).

1 Introduction

Law is one of the fields that may greatly benefit from the huge Artificial Intelligence (AI) advances, particularly in connection to language technologies. In fact, one can almost say that AI is changing the field. These changes are reflected in recently coined terms such as "Legal AI", which encompasses hundreds of methods proposed for information retrieval, text/knowledge mining, and Natural Language Processing (NLP). In the literature, NLP is often restricted to text processing, but we take the overarching view of covering both written and spoken language processing, as both text and speech processing are playing a vital role in shaping the future of legal AI.

I. Trancoso (✉) · N. Mamede · B. Martins · H. S. Pinto
INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal
e-mail: isabel.trancoso@inesc-id.pt; nuno.mamede@inesc-id.pt; bruno.g.martins@inesc-id.pt; sofia@inesc-id.pt

R. Ribeiro
INESC-ID and Iscte - Instituto Universitário de Lisboa, Lisboa, Portugal
e-mail: ricardo.ribeiro@inesc-id.pt

Thus, we structured this necessarily brief review into two main sections, covering text and speech analysis. We describe how the area has changed in the last decade, and how different language technologies may contribute to draft, dictate, analyse, and anonymise legal documents, streamline legal research, predict rulings, transcribe court proceedings, etc. Moreover, the chapter also attempts to draw attention to potential misuses of language technology, and their impact in the legal domain.

2 Language Processing Technologies for Processing Textual Data

Natural Language Processing (NLP) in the legal domain (Zhong et al. 2020) has addressed text analysis tasks such as legal judgment prediction (Aletras et al. 2016; Chen et al. 2019), legal topic classification (Chalkidis et al. 2021a), legal document retrieval and question answering, or contract understanding (Hendrycks et al. 2021), to name a few. As in other application areas for NLP, progress has often been made in connection to publicly available datasets, which researchers can use to evaluate system performance in a standardized way (e.g., the Legal General Language Understanding Evaluation benchmark is one recent example Chalkidis et al. 2021c). Joint evaluation initiatives (i.e., shared tasks) are also popular in the area. In these competitions, teams of researchers submit systems that address specific predefined challenges for the shared task, and the results are then evaluated against a "gold standard" that was previously prepared by the shared task organizers. Examples for shared tasks related to legal NLP include the Competition on Legal Information Extraction and Entailment (Rabelo et al. 2022), the Chinese AI and Law challenge, taking place yearly since 2018 (Zhong et al. 2018), or the Artificial Intelligence for Legal Assistance series of shared tasks, which started in 2019 (Bhattacharya et al. 2019). The field has also a long history, reflecting the changes that the general area of NLP has also seen over the years.

Up to the 1980s, most NLP systems were based on symbolic approaches leveraging hand-written rules. Starting in the late 1980s, there was a shift with the introduction of machine learning algorithms for NLP, using statistical inference to automatically learn *rules* through the analysis of large corpora. In the 2010s, representation learning and deep neural network-style machine learning methods became widespread in NLP. Popular techniques include the use of word embeddings to capture semantic properties of words, and an increase in end-to-end learning of higher-level tasks (e.g., question answering), instead of relying on a pipeline of separate intermediate tasks (e.g., parts-of-speech tagging and syntactic dependency parsing).

As with other specialized domains (e.g., biomedical or financial documents), legal text (e.g., legislation, court documents, contracts, etc.) has distinct characteristics compared to generic corpora, such as specialized vocabulary, a particular

syntax, semantics based on extensive domain-specific knowledge, or the common use of long sentences. These differences can affects the performance of generic NLP models, motivating research in this specific area. Even in the case of modern methods based on end-to-end learning, pre-training models with legal text can help to better capture the aforementioned characteristics, providing in-domain knowledge that is missing from other generic corpora.

In fact, several pre-trained legal language models, based on very large neural networks, have been recently introduced (Chalkidis et al. 2020b; Xiao et al. 2021). State-of-the-art NLP approaches are based on these types of models, following a design based on pre-training neural language models on huge amounts of (ideally in-domain) text, e.g. by considering unsupervised objectives such as predicting masked words from real sentences, followed by the supervised fine-tuning of these models to specific downstream tasks. The following sub-sections discuss different NLP applications related to the legal domain, often involving methods based on pre-trained neural language models.

2.1 Text Anonymization

Data anonymization is a process of masking or removing sensitive data from a document while preserving its original format. This process is important for sharing legal documents and court decisions without exposing any sensitive information (Mamede et al. 2016). Free-form text is a special type of document where data is contained in an unstructured way, as represented in natural language. Court decisions are examples of this type of document. From the content of these documents, it is necessary to identify text structures that represent names or unique identifiers, known as entities. This task is commonly referred to as NER (Named Entity Recognition). The three main classes of NEs are: person, location, and organization. Other important classes include dates, phone numbers, car plates, bank account references (eg. IBAN), and websites.

The main use of automatic text anonymization systems is to de-identify medical records and court decisions. A generic anonymization system is usually composed of up to four modules: (1) a module that normalizes the text and performs feature extraction; (2) a set of NE classifiers; (3) a poll to vote the most probable class of NE; and (4) a module that applies an anonymization method over the NEs and replaces the occurrences of these entities in the text.

One of the first automated anonymization systems was Scrub. It was introduced by Sweeney (1996) and it uses pattern-matching and dictionaries. The system runs multiple algorithms in parallel to detect different classes of entities. In 2006, part of the i2b2 (Informatics for Integrating Biology to the Bedside) Challenge was dedicated to the de-identification of clinical data. Seven systems participated in this challenge. The MITRE system, developed by Wellner et al. (2007), achieved the highest performance. The MITRE system uses two model-based NER tools, one based on Conditional Random Fields (CRF) and another on Hidden Markov Models.

Gardner and Xiong (2008) developed the Health Information DE-Identification (HIDE) framework for de-identification of private health information (PHI), which uses a NER tool based on CRF. Neamatullah et al. (2008) developed the MIT De-id package. This package is a dictionary and rule-based system and was made available for free on the Internet by PhysioNet. Uzuner et al. (2008) developed the Stat De-id that runs a set of classifiers in parallel. Each classifier is specialized in detecting a different category of entities. The Best-of-Breed System (BoB) by Ferránez et al. (2013), a hybrid design system, uses rules and dictionaries to score a higher recall, and it also uses model-based classifiers to score a higher precision.

Michaël Benesty[1] draws attention to the importance of the processing speed of anonymization systems. The case study was conducted in collaboration with the French administration and a French supreme court (*Cour de cassation*). More recently, Glaser et al. (2021b) presented a machine learning approach for the automatic identification of sensitive text elements in German legal court decisions. The adopted strategy includes several deep neural networks based on generic pre-trained contextual embeddings.

The most usual methods of anonymization include: suppression, tagging, random substitution, and generalization. The *suppression* method is a simple way of anonymizing a text that consists of the suppression of the NE using a neutral indicator that replaces the original text, e.g. 'XXXXXX'. The *tagging* method consists of the replacement of the NE with a label that could indicate its class and a unique identifier. It can be implemented by concatenating the class given by the NER tool and a unique numeric identifier, e.g. [**Organization123**]. The *random substitution* method replaces a NE with another random entity of the same class and morphosyntactic features. This method can be implemented using a *default list* containing random entities of each class. In highly inflected languages, it is important to replace entities of every class with another entity with the same gender and number. *Generalization* is any method of replacing an entity from the text with another entity that mentions an object of the same class but in a more general way, e.g.: *University of Lisbon* could be generalized to *University*, or even to *Institution*.

Some of the major problems of developing anonymization systems to be used in legal documents and court decisions are: (1) the lack of non-anonymized data sets, making impossible to compare approaches and making the evolution of these systems harder; (2) each jurisdiction features different distributions of named entity types and introduces court-specific anonymization challenges; (3) all the entities that refer to the same object within the document should be replaced by the same label, which implies the existence of a Co-reference resolution module which is also a big challenge; (4) The *random substitution* implies the extraction of the grammatical gender and number of the NE that is given by its headword. The headwords of NEs and their features must be determined at a pre-processing stage. Determining the

---

[1] https://towardsdatascience.com/why-we-switched-from-spacy-to-flair-to-anonymize-french-legal-cases-e7588566825f.

gender of the headword is important for replacing NEs that refer persons by another NE from the same gender, e.g. replacing *John* by *Peter*, or replacing *Mary* by *Anna*.

2.2 Document Classification

Text classification (i.e., the assignment of documents to classes from a pre-defined taxonomy) has many potential applications in the legal domain, particularly for categorizing legislative documents and cases. This can aid the process of legal research, and the development of knowledge management systems (Boella et al. 2016). Several studies have focused on legislative contents or court cases (Tuggener et al. 2020; De Araujo et al. 2020; Papaloukas et al. 2021), with some authors highlighting that legal document classification can be significantly harder than more generic text classification problems (Nallapati and Manning 2008).

In the specific case of legislative contents, much work on topic classification has focused on EU legislation documents, both in monolingual settings focusing on the English language (Chalkidis et al. 2019b, 2020a), and also on multi-lingual settings (Avram et al. 2021; Chalkidis et al. 2021a). These previous efforts addressed the task of classifying EU laws into EuroVoc[2] concepts, seeing the problem as a challenging instance of Large-scale Multi-label Text Classification (LMTC), given the need for assigning, to each given document, a subset of labels from a large predefined set (i.e., thousands of classes that are hierarchically organized), and given also the need for handling few and zero-shot scenarios (i.e., the distribution for how labels are assigned is highly skewed, and some labels have few or no training examples). A battery of state-of-the-art LMTC methods have been empirically evaluated, with very good results currently being obtained with approaches based on combining large pre-trained neural language models (i.e., models based on pre-trained Transformer-based approaches like BERT) with label-wise attention networks (i.e., using different parameters for weighting the document representations, according to each possible label). For instance, in experiments with 57k English legislative documents from EURLEX,[3] studies have reported values of 80.3 in terms of an R-Precision@K evaluation metric (Chalkidis et al. 2020a).

Document classification technology is also nowadays deployed in many practical settings. One interesting example is the JRC EuroVoc Indexer (JEX[4]), i.e. an open-source tool currently being used in many different settings, that was developed by the European Commission's Joint Research Centre (JRC) for automatically classifying documents according to EuroVoc descriptors, covering the 22 official EU languages. JEX can be used as a tool for interactive multi-label EuroVoc descriptor

---

[2] http://eurovoc.europa.eu/.

[3] http://eur-lex.europa.eu/.

[4] http://joint-research-centre.ec.europa.eu/language-technology-resources/jrc-eurovoc-indexer-jex_en.

assignment, which is particularly useful to increase the speed and consistency of human categorization processes, or it can be used fully automatically.

2.3 Information Retrieval

The need for handling large amount of digital documents has made the legal sector interesting for the development of specific methodologies for the management, storage, indexing, and retrieval of legal information. All these tasks fall into the realm of information retrieval, which manly focuses on information search problems where a description of the current situation (i.e., an information need) is used to query an automated system to retrieve the most suitable information, within a large repository, for the input query (Sansone and Sperlí 2022).

Work on legal information retrieval goes back to the 1960s (Wilson 1962; Eldridge and Dennis 1963; Choueka et al. 1971), but recent scientific developments are strongly connected to the Competition on Legal Information Extraction/Entailment (COLIEE), and to specific applications related to case law retrieval (Locke and Zuccon 2022). From 2015 to 2017 the COLIEE task was to retrieve Japanese Civil Code articles given a question, and since then the main COLIEE retrieval task has been to retrieve supporting cases given a short description of an unseen case. Most submitted systems leverage sparse representations of documents and queries, based on word occurrences, together with simple numerical statistics that reflect how important a word is to a document within a collection (i.e., statistics such as TF-IDF or BM25). More recent studies, both in the context of COLIEE and in other separate publications, have started to explore recent advances connected to neural ranking models (e.g., using large language models trained in text matching data, for re-ranking the results of simpler methods based on word-level statistics).

An interesting recent study has for instance focused on the task of regulatory information retrieval (Chalkidis et al. 2021b), which concerns retrieving all relevant laws that a given organization should comply with, or vice-versa (i.e., given a new law retrieve all the regulatory compliance controls, within an organization, that are affected by this law). Applications like this are much more challenging than traditional information retrieval tasks, where the query typically contains a few informative words and the documents are relatively short. In the case regulatory information retrieval (and also other legal tasks, such as similar case matching (Xiao et al. 2019)), the query is also a long document (e.g., a regulation) containing thousands of words, most of which are uninformative. Consequently, matching the query with other long documents, where the informative words are also sparse, becomes extremely difficult for traditional approaches based on word-level matching. Leveraging datasets composed of EU directives and UK regulations, which can serve both as queries and documents (i.e., a UK law is relevant to the EU directives it transposes, and vice versa), the authors reported on very good results (i.e., averaging over different queries, approximately 86.5% of the documents that are retrieved on the top 100 positions are relevant) with a system that combines

standard BM25 retrieval with result re-ranking through a neural language model fine-tuned to documents and specific tasks from the legal domain.

2.4   Information Extraction

Information extraction concerns automatically gathering and structuring important facts from textual documents (e.g., about specific types of events, or about entities and relationships between previously defined entities), facilitating the development of higher-level applications, in the sense that these can now focus on the analysis of structured information, as opposed to unstructured or semi-structured traditional legal texts. Several studies (Chalkidis et al. 2019c; Hendrycks et al. 2021) have explored information extraction from contracts, e.g. to extract information elements such as the contracting parties, agreed payment amount, start and end dates, or applicable law. Other studies focused on extracting information from legislation (Angelidis et al. 2018) or court cases (Leitner et al. 2019).

2.5   Summarization

A text summary conveys to the reader the most relevant content of one or more textual information sources, in a concise and comprehensible manner. The goal of a text summarization system is to automatically create such a document. This new document, the summary, is characterized by several aspects, such as the origin of the content, the number of input units, or the coverage of the summary. Regarding its content, a summary might be composed by extracts—*extractive summarization*—, directly taken from the input, or paraphrases—*abstractive summarization*—, which convey the content of a passage of the input using a different wording. In relation to the number of input units, if the input consists of only one document, the task is designated *single-document summarization*, when dealing with several input documents, the problem is named as *multi-document summarization*. Finally, concerning the coverage of the input source(s), it can be comprehensive, when creating generic summaries, or selective, if driven by an input query.

Several difficulties arise when addressing this task, but one of utmost importance is how to assess the relevant content. Different methods have been explored since the first experiments reported by Luhn (1958) and Edmundson (1969). They established an important research direction, which can be named as feature-based passage scoring, which explored features based on term weighting, sentence position, sentence length, or linguistic information. In the 2000s, centrality-based methods (Radev et al. 2004; Erkan and Radev 2004; Zhu et al. 2007; Kurland and Lee 2010; Ribeiro and de Matos 2011), independently of the underlying representation, attracted much attention. Geometric centroids, graph-based ranking methods, or, in general, representations in which "recommended" passages were the focus of

this line of work. All of these were unsupervised approaches, in which we can also include important methods such as Maximal Marginal Relevance (Carbonell and Goldstein 1998) or Latent Semantic Analysis-based (Gong and Liu 2001) methods. In addition to these kind of approaches, some supervised methods were also explored (Wong et al. 2008). Recently, most of the research on this topic has been based on neural networks, with sequence-to-sequence models attracting a significant amount of attention (Rush et al. 2015; See et al. 2017; Celikyilmaz et al. 2018), as well as work using pre-trained language models as encoders (Liu and Lapata 2019; Manakul and Gales 2021).

In what concerns legal document summarization, some of the oldest work goes back to 2004. Farzindar and Lapalme (2004) present an approach to generate short summaries of records of the proceedings of federal courts in Canada. This work focused on extracting the most important textual units, following a feature-based passage scoring approach. The pipeline includes a thematic segmentation specifically directed at legal documents, aimed at discovering the structure of the judgment record. This segmentation is followed by filtering and selection stages. The former filters out citations and other noisy content and the latter extracts textual units based on their score, computed using features such as the position of the paragraphs in the document, the position of the paragraphs in the thematic segment, the position of the sentences in the paragraph, the distribution of the words in the document, and TF-IDF. Closely related to this work, is the system proposed by Hachey and Grover (2006), in the sense that these authors also develop a classifier for rhetorical information. The extraction of relevant sentences is also cast as a classification problem, following a supervised machine learning-based approach. Features such as location, thematic words, sentence length, quotation level, entities, and cue phases are explored in both a Naïve Bayes classifier and a Maximum Entropy classifier.

More recent work in legal document summarization followed the same trend we observe in generic automatic text summarization and natural language processing in general, which means that it is based on neural networks and pre-trained language models. The work reported by Glaser et al. (2021a) focuses on German court rulings. The system includes a dedicated pre-processing step where norms, anonymization tokens, and references to other legal documents, for instance, are addressed. Concerning the summarization method, the authors explored both extractive and abstractive approaches. Word are encoded using GloVe embeddings (Pennington et al. 2014) and for sentence representation three approaches are explored: CNN, GRU, or attention. The final sentence representation is given by a cross-sentence CNN or RNN that captures information from neighboring sentences. The selection score is given by a sigmoid function. The abstractive approach is similar, but instead of using a cross-sentence CNN or RNN, the sentence embeddings are aggregated using a RNN, creating an embedding for the document. The approach follows an encoder-decoder structure. As expected, the abstractive approaches had a worst performance when compared to extractive approaches and baselines, even considering older, centrality-based, approaches such as LexRank. The proposed

extractive approaches achieved the best results showing that neural networks-based approaches are also adequate for specific domains.

An interesting work based on pre-trained language models is the system proposed in Savelka and Ashley (2021). Their work is closely related to summarization, but instead they focus on the idea of how well a sentence explains a legal concept, based on data from the Caselaw access project (U.S.A. legal cases from different types of courts). The selected data was human classified into four categories: high value, certain value, potential value, and no value. They fine-tuned RoBERTa (Liu et al. 2019), a model derived from the well-known BERT (Devlin et al. 2019), as the base for their experiments. They explore three approaches: one that predicts the class using only the input sentence; the second approach uses the pair legal concept-sentence; and, finally, the last variation is based on the pair composed by "the whole provision of written law" and the sentence. One important conclusion is that, given the success of the first experiment, sentences carry information about their usefulness, which is strongly related to understanding if a sentence is a good candidate to be in a summary.

2.6   Question Answering and Conversational Systems

Legal question answering concerns the retrieval and analysis of information within knowledge repositories (e.g., large document collections), so as to provide accurate answers to legal questions. Typical users for legal question answering systems can include litigators seeking answers to case-specific legal questions (Khazaeli et al. 2021), or laypersons seeking to better understand their legal rights (Ravichander et al. 2019).

The task requires identifying relevant legislation, case law, or other legal documents, and extracting the elements within those documents that answer a particular question. In some cases, the extracted information elements also need to be further summarized into a concise answer. As evident from the previous problem definition, legal question answering typically involves combining techniques from information retrieval and extraction, and the aforementioned Competition on Legal Information Extraction and Entailment has also been a notable venue for reporting advances in this domain.

On what regards industrial applications, it is interesting to note that companies such as IBM Watson Legal, LegalMation, or Ross Intelligence have developed custom question answering commercial products based on Watson, i.e. a question answering system developed by IBM that, in 2011, won the Jeopardy challenge against the TV quiz show's two biggest all-time champions (Ferrucci et al. 2010). Watson's architecture features a variety of NLP technologies, including parsing, question classification, question decomposition, automatic source acquisition and evaluation, entity and relation detection, logical form generation, and knowledge representation and reasoning. After the success in Jeopardy, the system was reorga-

nized and commercialized, by combining and customizing its specific modules to specific domains and tasks (i.e., not just question answering, but also other tasks).

2.7 Predictions Supported on Textual Evidence

The idea that computers can predict the outcome of legal cases goes back to the early 60s, where Lawlor (1963) considers "the analysis and prediction of judicial decisions" one of the most important tasks to which computer technology can contribute. However, most of the work on the task is not that old. In fact, we can understand the difficulty of the endeavor as even recent literature overviews (e.g., Robaldo et al. 2019) focusing on natural language understanding in the legal domain address topics like resource construction or simple information extraction tasks.

In 2004, Ruger et al. (2004) compare the performance of classification trees against a group of legal experts to predict the outcome of the 2002 Term of the United States Supreme Court. The used features were the following: the provenience of the case until it reached the Supreme Court, the issue area of the case, the type of petitioner, the type of respondent, the lower court ruling political orientation (liberal or conservative), and the claim of unconstitutionality of a law or practice by the petitioner. The results were surprising: the automatic method correctly predicted 75% of the Court's affirm/reverse outcomes, while the experts accuracy was only of 59.1%. The best results were achieved for economic activity cases and the worst for federalism, where the accuracy for both the experts and the automatic approach was similar.

A different two-step approach was followed by Brüninghaus and Ashley (2005), focusing on textual information. The first step extracts a representation of the legal case, based on its words (a bag-of-words representation obtained by removing punctuation, numbers and stop words); named entity recognition (where names and case specific instances are replaced by their type); and syntactic relations. This representation is submitted to a set of classifiers that capture several aspects that are used to represent the cases, designated as Factors (e.g., *Agreed-Not-To-Disclose*, *Security-Measures*, or *Agreement-Not-Specific*). The used data consisted in the Trade Secret Law knowledge base, which includes 146 cases from the CATO system, an intelligent learning environment for new students that begin studying law, already represented in terms of Factors (Alaven 1997). The second step predicts the legal outcome based on the Factors representation, using case-based reasoning (Bruninghaus and Ashley 2003).

Focusing in the European Court of Human Rights and exploring only textual content, Aletras et al. (2016) experiment with Support Vector Machines classifiers (linear kernel) based on n-gram and topic-based representations. The goal is to predict if a certain case violates a specific article of the Convention, achieving accuracy rates close 80% in this binary classification problem. Şulea et al. (2017) also use Support Vector Machines classifiers (linear kernel), but concentrate on the French Supreme Court (*Cour de Cassation*) rulings. The authors address three tasks:

predicting the law area of a case, predicting the court ruling, and estimating when a case description and a ruling were issued. Preprocessing included the removal of diacritics and punctuation and the lowercasing of all the words. Features consisted of unigrams and bigrams. In what concerns predicting the court ruling, two variations of the task considering six (first-word ruling) and eight (full ruling) classes were addressed. Results were promising as in both experiments the proposed approach achieved f1 scores and accuracies over 90%.

As previously mentioned, most of the work on this topic is recent and thus explores deep neural network models. Chalkidis et al. (2019a) explore different network architectures—a bidirectional Gated Recurrent Units with self-attention-based architecture, a hierarchical attention network, a label-wise attention network, and two BERT-based architectures (a regular and a hierarchical version)—to predict the violation of articles of the European Court of Human Rights. Differently from Aletras et al. (2016), they do not restrict to specific articles, additionally exploring a multi-label classification view of the task. The simple BERT-based approach was the poorest performing architecture, with the best results being achieved by hierarchical architectures. Confirming the importance of hierarchical approaches, Zhu et al. (2020) also explore a hierarchical attention network-based architecture in legal judgment prediction in the context of criminal cases published by the Chinese government from China Judgment Online. Alghazzawi et al. (2022) combine a long short-term memory with a convolutional neural network to address the same problem as Ruger et al. (2004), i.e., to predict the affirm/reverse/other outcome of the US Supreme Court rulings.

Finally, Medvedeva et al. (2022) provide an interesting overview of this topic, while addressing relevant concepts clarification. The authors argue that a clear and well-defined terminology is important for the advancement of the research in this topic, namely distinguishing between three different tasks: outcome identification, outcome-based judgment categorization, and outcome forecasting.

2.8   Summary

This section described the development status of language technologies targeting various legal text processing tasks. Although our survey has mostly focused on recent academic developments, a large number of companies, including hundreds of start-ups, are also currently operating in the emerging "Legal AI" industry, providing text analytics services that target a wide variety of use cases that are currently poorly handled, e.g. due to the excessive amount of data (i.e., documents) that poses challenges for human analysis. In all the surveyed tasks, recent developments associated to deep learning methods (e.g., pre-trained neural language models specifically targeting legal text) have brought forward significant improvements. Current challenges in the area relate, for instance, to the combination of deep neural networks with knowledge-based methods (e.g., to improve interpretability and to better account with expert knowledge and legal reasoning), or to techniques enabling

a better control of potential biases in model results (e.g., gender biases or racial discrimination).

3   Spoken Language Technologies

The use of spoken language technologies in the legal domain is also becoming increasingly pervasive, mostly because of the significant increase in performance achieved by deep learning techniques. This justifies a review of this recent progress and its impact in the legal domain.

3.1   *Automatic Speech Recognition*

The state of the art in automatic speech recognition (ASR) before the advent of deep learning was predominantly based on the GMM-HMM paradigm (Gaussian Mixture Models - Hidden Markov Models). By feeding these acoustic models with perceptually meaningful features, and combining them with additional knowledge sources provided by n-gram language models and lexical (or pronunciation) models, one achieved word error rates (WER) that made ASR systems usable for certain tasks. Dictation was one of these tasks, most particularly for the legal and healthcare domains (e.g. radiology reports), characterized by clean recording conditions and relatively formal documents. Acoustic models could be adapted to the speaker, and lexical and language models could be adapted to the domain, allowing the use of ASR by lawyers for dictating documents, case notes, briefs, contracts and correspondence. However, the uptake of such applications was not significant and depended heavily on the availability of resources to train models for different languages/accents.

For nearly three decades, progress was relatively stale for ASR, until the emergence of the so-called 'hybrid paradigm', that pairs deep neural networks with HMMs. Nowadays, models trained with nearly 1000 h of read audiobooks achieve a WER of 3.8%, an unthinkable result a decade ago. As in many other AI domains, fully end-to-end architectures have also been proposed to perform the entire ASR pipeline (Karita et al. 2019), with the exception of feature extraction, but the performance is significantly worse when training data is short. In fact, *There is no data like more data* is a citation from an ASR researcher back in the 80s, which is still valid nowadays, representing a huge challenge when porting to another domain.

Transcribing audiobooks or dictating legal documents are relatively easy tasks for ASR systems. Their application to conversational speech is much more challenging, with error rates that are almost triple the above results. The presence of other factors such as non-native accents, or distant microphones in a meeting room, may also have a very negative impact.

All these challenges motivate the use of a panoplia of machine learning approaches: audio augmentation (Park et al. 2019; Ko et al. 2015), transfer learning (Abad et al. 2020), multi-task learning (Pironkov et al. 2016), etc. Also worth mentioning are the recent unsupervised approaches that leverage speech representations to segment unlabeled audio and learn a mapping from these representations to phonemes via adversarial training (Baevski et al. 2021).

This enormous boost in ASR performance has had a great impact in the legal domain. For dictation in this domain where training data has been increasingly amassed, several companies now claim WERs below 1%, announcing streamlined ways to dictate documents, 3–5 times faster than typing, altogether reducing liability and compliance. This progress was particularly relevant during the recent pandemic which has forced more lawyers to work remotely, without an easy access to pools of typists.

But the use of ASR in the legal domain is not at all restricted to the dictation task. Transcribing video and audio evidence into legal transcripts is an increasingly essential task for presenting in court, for making appeals, etc. The manual process can take up more than 5 times real time, which is a significant motivation for speeding it up by correcting an automatic transcript instead of transcribing from scratch. Transcribing audio court proceedings is another increasingly common use of ASR in the legal domain. Spotting keywords in tapped conversations may also be particularly relevant for intelligence services.

Besides the above-mentioned challenges of recognizing spontaneous speech, all these types of transcript require a previous task of speaker diarization - recognizing who spoke when (Tumminia et al. 2021). This task is closely related to automatic speaker recognition and may be particularly complex in scenarios where speaker overlap is frequent.

Due to their high complexity, ASR systems typically run in the cloud. In personal voice assistants, the task of spotting the wake up keyword when they are in the "always listening mode" is done on device using much less complex approaches. There is a growing public awareness of the privacy concerns over this "always listening mode". Requests for these recordings by suspects who would like to present them as proof in court have been so far denied by companies such as Amazon.[5]

3.2 *Speaker Recognition and Speaker Profiling*

The importance of identifying speakers in recordings has been realised by law enforcement agencies and intelligence services who used to rely in experts to manually analyse the so-called 'voice prints', long before automatic speaker

---

[5] https://techcrunch.com/2016/12/27/an-amazon-echo-may-be-the-key-to-solving-a-murder-case/.

recognition systems reached the performance levels that allowed their use in such a domain. Much of this recent progress may be attributed to representation learning, the so-called 'speaker embeddings', which encode the speaker characteristics of an utterance of variable duration into a fixed-length vector. The most popular technique for achieving this compact representation is currently the x-vector approach (Snyder et al. 2016). These embeddings are extracted from the hidden layers of deep neural networks, when they are trained to distinguish over thousands of speakers. In fact, this approach has been applied to Voxceleb,[6] a multimodal corpus of YouTube clips that includes over 7000 speakers of multiple ethnicities, accents, occupations and age groups, reaching impressive equal error rates (EER) close to 3%. This metric derives its name from corresponding to a threshold for which the false positive and false negative error rates are equal. Confidence measures may be particularly important in the legal context.

The area of speaker profiling is one of the most recent ones in speech processing. Speech is a biometric signal that reveals, in addition to the meaning of words, much information about the user, including his/her preferences, personality traits, mood, health, and political opinions, among other data such as gender, age range, height, accent, etc. Moreover, the input signal can be also used to extract relevant information about the user environment, namely background sounds. Powerful machine learning classifiers can be trained to automatically detect speaker traits that may be of particular importance to law enforcement agencies and intelligence services.

These endless possibilities for profiling speakers from their voices raise many privacy concerns about the misuse of such technologies.

3.3   Speech Synthesis and Voice Conversion

A decade ago, the state of the art in text-to-speech (TTS) was dominated by concatenative techniques that selected the best segments to join together from a huge corpus of sentences read by a single speaker. The concatenative synthesis module was typically preceded by a complex chain of linguistic processing modules which took text as input and produced a string of phonemes together with the prosodic information that specified the derived intonation. Despite significant improvements, namely through the use of hybrid approaches (Qian et al. 2013) that combined statistical parametric and concatenative techniques, the synthetic speech quality was still very different in naturalness from human speech, expressiveness was very limited, and the costs of building new synthetic voices were often prohibitive.

A major breakthrough was achieved in the mid 2010s by replacing the traditional concatenative synthesis module by a deep neural network module that took as input time-frequency spectrogram representations (van den Oord et al. 2016).

---

[6] https://www.robots.ox.ac.uk/~vgg/data/voxceleb/.

Later, the whole paradigm changed to encoder-decoder architectures, with attention mechanisms mapping the linguistic time scale to the acoustic time scale (Shen et al. 2018).

These advances led to multi-speaker TTS systems leveraging speaker embeddings, and opened the possibility of building synthetic voices with only a few seconds of a new voice, using for instance, flow-based models (Kim et al. 2020; Casanova et al. 2021). The synthetic speech quality became very close to human speech, reaching values above 4 on a scale of 1–5, the so-called 'Mean Opinion Score' (MOS) scale.

The possibility of disentangling linguistic contents and speaker embeddings was indeed crucial not only for text-to-speech systems but for voice conversion (VC) systems as well. In VC, the input is speech instead of text and the goal may be changing the voice identity, the emotion, the accent, etc.

This disentanglement can be achieved using, for instance, variational auto-encoder schemes, in which the linguistic content encoder learns a latent code from the source speaker speech, and the speaker encoder learns the speaker embedding from the target speaker speech. At run-time, the latent code and the speaker embedding are combined to generate speech in the voice of the target speaker. Nowadays, new approaches try to factor in prosody embeddings or style embeddings as well.

Moreover, the artificial voice may not correspond to a target speaker but, for instance, to an average of a set speaker embeddings selected among the ones farthest from the original speaker embedding. This is in fact one of the many approaches proposed for speaker anonymization.

The repercussions of this progress in TTS/VC in the legal framework are potentially huge since it can be used for the purposes of incrimination, defamation or misinformation. Detecting a *deep fake* voice from an original voice will be increasingly more difficult and one may wonder when audio evidence will no longer be admissible in court. At the same time, impersonating speakers may lead to crimes that were not feasible with the technologies we had a decade ago. In fact major thefts have already been reported.[7]

On the other hand, synthetic voices may be used for attacking (or spoofing) automatic speaker verification systems. In fact, as the quality of TTS/VC with very little spoken material from a target speaker increases, the need for more sophisticated anti-spoofing also grows concurrently. Last but not least, one should also mention the possibility of hidden voice commands, injected in the input signal. This threat has always existed, namely through exploiting the fact that the human hear could not detect certain signals. But nowadays adversarial attacks raise this threat to a new level, making us aware of how vulnerable deep learning techniques may be to the perturbation of a classifier's input at test time such that the classifier outputs a wrong prediction. In the past, such perturbation was in

---

[7] https://www.washingtonpost.com/technology/2019/09/04/an-artificial-intelligence-first-voice-mimicking-software-reportedly-used-major-theft/.

most cases perceptible, but with adversarial attacks one can now generate highly imperceptible perturbations that are extremely effective in misleading either speaker or speech recognition systems. Such techniques are just one example of the endless possibilities for misuse of the AI driven speech technologies. We have barely touched the surface, in terms of attacks that may target speech-based apps.

4   Conclusions

This chapter tried to do give a very condensed overview of language technologies for the legal domain, running the risk of very soon becoming outdated, such is the tremendous progress in the field nowadays. The brief overview of these recent advances may be misleading, giving the impression that embedding-based methods will solve all classes of problems whether for written or spoken language processing in the legal domain, provided there are large-enough training datasets. In fact, many researchers are working on machine learning alternatives (e.g. zero-shot or few-shot learning) to cope with tasks for which such datasets are not available. However, combining embedded-based approaches with symbol-based methods remains a challenge that may significantly contribute to greater interpretability.

Another challenge to be addressed by the forthcoming generation of AI legal tools is keeping track of changing regulations by propagating the consequences of these changes on those issues that depend on them, since it requires a smooth integration of embedded-based and symbol-based approaches.

With progress also comes a greater awareness of the ethical issues of language technologies in the legal domain. In particular, in what concerns gender bias and racial discrimination, which may be extremely important for tasks such as judgement prediction.

We have also alerted to the potential misuse of speech technologies for impersonation or spoofing and last but not least to the privacy issues that are involved in the remote processing of a signal such as speech that must be legally regarded as PII (Personally Identifiable Information).

References

Abad A, Bell P, Carmantini A, Renals S (2020) Cross lingual transfer learning for zero-resource domain adaptation. In: 2020 IEEE International conference on acoustics, speech and signal processing (ICASSP), pp 6909–6913. https://doi.org/10.1109/ICASSP40776.2020.9054468

Alaven V (1997) Teaching case-based argumentation through a model and examples. Ph.D Thesis, University of Pittsburgh

Aletras N, Tsarapatsanis D, Preoţiuc-Pietro D, Lampos V (2016) Predicting judicial decisions of the European court of human rights: A natural language processing perspective. PeerJ Comput Sci 2:e93

Alghazzawi D, Bamasag O, Albeshri A, Sana I, Ullah H, Asghar MZ (2022) Efficient prediction of court judgments using an LSTM+CNN neural network model with an optimal feature set. Mathematics 10(5). https://doi.org/10.3390/math10050683. https://www.mdpi.com/2227-7390/10/5/683

Angelidis I, Chalkidis I, Koubarakis M (2018) Named entity recognition, linking and generation for Greek legislation. In: JURIX, pp 1–10

Avram AM, Păis V, Tufis DI (2021) PyEuroVoc: a tool for multilingual legal document classification with EuroVoc descriptors. In: Proceedings of the international conference on recent advances in natural language processing (RANLP 2021), pp 92–101

Baevski A, Hsu WN, Conneau A, Auli M (2021) Unsupervised speech recognition. Preprint, ArXiv 2105.11084

Bhattacharya P, Ghosh K, Ghosh S, Pal A, Mehta P, Bhattacharya A, Majumder P (2019) FIRE 2019 AILA track: artificial intelligence for legal assistance. In: Proceedings of the 11th forum for information retrieval evaluation. Association for Computing Machinery, New York, NY, USA, FIRE '19, pp 4–6. https://doi.org/10.1145/3368567.3368587

Boella G, Caro LD, Humphreys L, Robaldo L, Rossi P, van der Torre L (2016) Eunomos, a legal document and knowledge management system for the web to provide relevant, reliable and up-to-date information on the law. Artif. Intell. Law 24(3):245–283

Bruninghaus S, Ashley KD (2003) Predicting outcomes of case based legal arguments. In: Proceedings of the 9th international conference on artificial intelligence and law. Association for Computing Machinery, ICAIL '03, pp 233–242. https://doi.org/10.1145/1047788.1047838

Brüninghaus S, Ashley KD (2005) Generating legal arguments and predictions from case texts. In: Sartor G (ed) The tenth international conference on artificial intelligence and law, proceedings of the conference, June 6–11, 2005, Bologna, Italy, ACM, pp 65–74. https://doi.org/10.1145/1165485.1165497

Carbonell JG, Goldstein J (1998) The use of MMR, diversity-based reranking for reordering documents and producing summaries. In: Croft WB, Moffat A, van Rijsbergen CJ, Wilkinson R, Zobel J (eds) SIGIR '98: proceedings of the 21st annual international ACM SIGIR conference on research and development in information retrieval, August 24–28 1998, Melbourne, Australia, ACM, pp 335–336. https://doi.org/10.1145/290941.291025

Casanova E, Shulby C, Gölge E, Müller NM, de Oliveira FS, Candido Jr A, da Silva Soares A, Aluisio SM, Ponti MA (2021) SC-GlowTTS: an efficient zero-shot multi-speaker text-to-speech model. In: Interspeech, pp 3645–3649. https://doi.org/10.21437/Interspeech.2021-1774

Celikyilmaz A, Bosselut A, He X, Choi Y (2018) Deep communicating agents for abstractive summarization. In: Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: human language technologies. Association for computational linguistics, New Orleans, Louisiana, vol 1 (Long Papers), pp 1662–1675. https://doi.org/10.18653/v1/N18-1150. https://aclanthology.org/N18-1150

Chalkidis I, Androutsopoulos I, Aletras N (2019a) Neural legal judgment prediction in English. In: Proceedings of the 57th annual meeting of the association for computational linguistics, association for computational linguistics, Florence, Italy, pp 4317–4323. https://doi.org/10.18653/v1/P19-1424. https://aclanthology.org/P19-1424

Chalkidis I, Fergadiotis E, Malakasiotis P, Androutsopoulos I (2019b) Large-scale multi-label text classification on EU legislation. In: Proceedings of the annual meeting of the association for computational linguistics, pp 6314–6322

Chalkidis I, Fergadiotis M, Malakasiotis P, Androutsopoulos I (2019c) Neural contract element extraction revisited. In: Workshop on document intelligence at NeurIPS 2019

Chalkidis I, Fergadiotis M, Kotitsas S, Malakasiotis P, Aletras N, Androutsopoulos I (2020a) An empirical study on large-scale multi-label text classification including few and zero-shot labels. Preprint, arXiv:201001653

Chalkidis I, Fergadiotis M, Malakasiotis P, Aletras N, Androutsopoulos I (2020b) LEGAL-BERT: The muppets straight out of law school. In: Findings of the association for computational linguistics: EMNLP 2020, pp 2898–2904

Chalkidis I, Fergadiotis M, Androutsopoulos I (2021a) Multieurlex-a multi-lingual and multi-label legal document classification dataset for zero-shot cross-lingual transfer. In: Proceedings of the conference on empirical methods in natural language processing, pp 6974–6996

Chalkidis I, Fergadiotis M, Manginas N, Katakalou E, Malakasiotis P (2021b) Regulatory compliance through Doc2Doc information retrieval: A case study in EU/UK legislation where text similarity has limitations. Preprint, arXiv:210110726

Chalkidis I, Jana A, Hartung D, Bommarito M, Androutsopoulos I, Katz DM, Aletras N (2021c) Lexglue: a benchmark dataset for legal language understanding in English. Preprint, arXiv:211000976

Chen H, Cai D, Dai W, Dai Z, Ding Y (2019) Charge-based prison term prediction with deep gating network. In: Proceedings of the conference on empirical methods in natural language processing and the international joint conference on natural language processing, pp 6362–6367

Choueka Y, Cohen M, Dueck J, Fraenkel AS, Slae M (1971) Full text document retrieval: Hebrew legal texts (report on the first phase of the responsa retrieval project). In: Proceedings of the 1971 international ACM SIGIR conference on information storage and retrieval. Association for Computing Machinery, pp 61–79. https://doi.org/10.1145/511285.511293

De Araujo PHL, de Campos TE, Braz FA, da Silva NC (2020) VICTOR: a dataset for Brazilian legal documents classification. In: Proceedings of the language resources and evaluation conference, pp 1449–1458

Devlin J, Chang MW, Lee K, Toutanova K (2019) 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. Association for Computational Linguistics, Minneapolis, Minnesota, vol 1 (Long and Short Papers), pp 4171–4186. https://doi.org/10.18653/v1/N19-1423. https://aclanthology.org/N19-1423

Edmundson HP (1969) New methods in automatic extracting. J ACM 16(2):264–285. https://doi.org/10.1145/321510.321519

Eldridge WB, Dennis SF (1963) The computer as a tool for legal research. Law Contemp. Probl. 28:78–99

Erkan G, Radev DR (2004) Lexrank: graph-based lexical centrality as salience in text summarization. J Artif Intell Res 22:457–479. https://doi.org/10.1613/jair.1523

Farzindar A, Lapalme G (2004) Legal text summarization by exploration of the thematic structure and argumentative roles. In: Text summarization branches out, association for computational linguistics, Barcelona, Spain, pp 27–34. https://aclanthology.org/W04-1006

Ferrández O, South B, Shen S, Friedlin F, Samore M, Meystre S (2013) BoB, a best-of-breed automated text de-identification system for VHA clinical documents. J Am Med Inform Assoc 20(1):77–83

Ferrucci DA, Brown EW, Chu-Carroll J, Fan J, Gondek D, Kalyanpur A, Lally A, Murdock JW, Nyberg E, Prager JM, Schlaefer N, Welty CA (2010) Building Watson: an overview of the DeepQA project. AI Mag 31(3):59–79. https://doi.org/10.1609/aimag.v31i3.2303

Gardner J, Xiong L (2008) HIDE: an integrated system for Health information DE-identification. In: Computer-Based Medical Systems. IEEE Computer Society, Washington, pp 254–259

Glaser I, Moser S, Matthes F (2021a) Summarization of German court rulings. In: Proceedings of the natural legal language processing workshop 2021, association for computational linguistics, Punta Cana, Dominican Republic, pp 180–189. https://doi.org/10.18653/v1/2021.nllp-1.19. https://aclanthology.org/2021.nllp-1.19

Glaser I, Schamberger T, Matthes F (2021b) Anonymization of German legal court rulings. Association for Computing Machinery, New York, pp 205–209. https://doi.org/10.1145/3462757.3466087

Gong Y, Liu X (2001) Generic text summarization using relevance measure and latent semantic analysis. In: Croft WB, Harper DJ, Kraft DH, Zobel J (eds) SIGIR 2001: Proceedings of the 24th annual international ACM SIGIR conference on research and development in information retrieval, September 9–13, 2001, New Orleans, Louisiana, USA, ACM, pp 19–25. https://doi.org/10.1145/383952.383955. https://doi.org/10.1145/383952.383955

Hachey B, Grover C (2006) Extractive summarisation of legal texts. Artif Intell Law 14(4):305–345. https://doi.org/10.1007/s10506-007-9039-z

Hendrycks D, Burns C, Chen A, Ball S (2021) CUAD: an expert-annotated NLP dataset for legal contract review. Preprint, arXiv:210306268

Karita S, Wang X, Watanabe S, Yoshimura T, Zhang W, Chen N, Hayashi T, Hori T, Inaguma H, Jiang Z, Someki M, Enrique N, Soplin Y, Yamamoto R (2019) A comparative study on transformer vs RNN in speech applications. In: IEEE automatic speech recognition and understanding workshop (ASRU), pp 449–456

Khazaeli S, Punuru J, Morris C, Sharma S, Staub B, Cole M, Chiu-Webster S, Sakalley D (2021) A free format legal question answering system. In: Proceedings of the natural legal language processing workshop 2021, pp 107–113

Kim J, Kim S, Kong J, Yoon S (2020) Glow-TTS: a generative flow for text-to-speech via monotonic alignment search. Preprint, ArXiv:2005.11129

Ko T, Peddinti V, Povey D, Khudanpur S (2015) Audio augmentation for speech recognition. In: Interspeech, pp 3586–3589. https://doi.org/10.21437/Interspeech.2015-711

Kurland O, Lee L (2010) Pagerank without hyperlinks: structural reranking using links induced by language models. ACM Trans Inf Syst 28(4):18:1–18:38. https://doi.org/10.1145/1852102.1852104

Lawlor RC (1963) What computers can do: analysis and prediction of judicial decisions. Am Bar Assoc J 49(4):337–344. http://www.jstor.org/stable/25722338

Leitner E, Rehm G, Moreno-Schneider J (2019) Fine-grained named entity recognition in legal documents. In: International conference on semantic systems. Springer, pp 272–287

Liu Y, Lapata M (2019) Text summarization with pretrained encoders. In: Inui K, Jiang J, Ng V, Wan X (eds) Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3–7, 2019. Association for Computational Linguistics, pp 3728–3738. https://doi.org/10.18653/v1/D19-1387

Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V (2019) Roberta: A robustly optimized BERT pretraining approach. CoRR abs/1907.11692. http://arxiv.org/abs/1907.11692

Locke D, Zuccon G (2022) Case law retrieval: problems, methods, challenges and evaluations in the last 20 years. Preprint, arXiv:220207209

Luhn HP (1958) The automatic creation of literature abstracts. IBM J Res Dev 2(2):159–165. https://doi.org/10.1147/rd.22.0159

Mamede N, Baptista J, Dias F (2016) Automated anonymization of text documents. In: 2016 IEEE congress on evolutionary computation (CEC), pp 1287–1294. https://doi.org/10.1109/CEC.2016.7743936

Manakul P, Gales MJF (2021) Long-span summarization via local attention and content selection. In: Zong C, Xia F, Li W, Navigli R (eds) Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing, ACL/IJCNLP 2021, vol 1 (Long Papers), Virtual Event, August 1–6, 2021. Association for Computational Linguistics, pp 6026–6041. https://doi.org/10.18653/v1/2021.acl-long.470

Medvedeva M, Wieling M, Vols M (2022) Rethinking the field of automatic prediction of court decisions. Artif. Intell. Law. https://doi.org/10.1007/s10506-021-09306-3

Nallapati R, Manning CD (2008) Legal docket classification: where machine learning stumbles. In: Proceedings of the 2008 conference on empirical methods in natural language processing, pp 438–446

Neamatullah I, Douglass M, Lehman L, Reisner A, Villarroel M, Long W, Szolovits P, Moody G, Mark R, Clifford G (2008) Automated de-identification of free-text medical records. BMC Medical Inform Decis Mak 8(1):1–17

van den Oord A, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior AW, Kavukcuoglu K (2016) WaveNet: a generative model for raw audio. CoRR abs/1609.03499

Papaloukas C, Chalkidis I, Athinaios K, Pantazi DA, Koubarakis M (2021) Multi-granular legal topic classification on Greek legislation. Preprint, arXiv:210915298

Park DS, Chan W, Zhang Y, Chiu CC, Zoph B, Cubuk ED, Le QV (2019) SpecAugment: a simple data augmentation method for automatic speech recognition. In: Interspeech, pp 2613–2617. https://doi.org/10.21437/Interspeech.2019-2680

Pennington J, Socher R, Manning CD (2014) Glove: global vectors for word representation. In: Moschitti A, Pang B, Daelemans W (eds) Proceedings of the 2014 conference on empirical methods in natural language processing, EMNLP 2014, October 25–29, 2014, Doha, Qatar. A meeting of SIGDAT, a Special Interest Group of the ACL, ACL, pp 1532–1543. https://doi.org/10.3115/v1/d14-1162

Pironkov G, Dupont S, Dutoit T (2016) Multi-task learning for speech recognition: an overview. In: ESANN - European symposium on artificial neural networks, computational intelligence and machine learning (ESANN), pp 189–194

Qian Y, Soong FK, Yan ZJ (2013) A unified trajectory tiling approach to high quality speech rendering. IEEE Trans Audio Speech Language Process 21(2):280–290. https://doi.org/10.1109/TASL.2012.2221460

Rabelo J, Goebel R, Kim MY, Kano Y, Yoshioka M, Satoh K (2022) Overview and discussion of the competition on legal information extraction/entailment (COLIEE) 2021. Rev Socionetwork Strategies 16, 111–133

Radev DR, Jing H, Sty M, Tam D (2004) Centroid-based summarization of multiple documents. Inf Process Manag 40(6):919–938. https://doi.org/10.1016/j.ipm.2003.10.006

Ravichander A, Black AW, Wilson S, Norton T, Sadeh N (2019) Question answering for privacy policies: Combining computational and legal perspectives. Preprint, arXiv:191100841

Ribeiro R, de Matos DM (2011) Centrality-as-relevance: Support sets and similarity as geometric proximity. J Artif Intell Res 42:275–308, https://doi.org/10.1613/jair.3387

Robaldo L, Villata S, Wyner A, Grabmair M (2019) Introduction for artificial intelligence and law: special issue "natural language processing for legal texts". Artif Intell Law 27(2):113–115. https://doi.org/10.1007/s10506-019-09251-2

Ruger TW, Kim PT, Martin AD, Quinn KM (2004) The supreme court forecasting project: legal and political science approaches to predicting supreme court decisionmaking. Columbia Law Rev 104(4):1150–1210

Rush AM, Chopra S, Weston J (2015) A neural attention model for abstractive sentence summarization. In: Màrquez L, Callison-Burch C, Su J, Pighin D, Marton Y (eds) Proceedings of the 2015 conference on empirical methods in natural language processing, EMNLP 2015, Lisbon, Portugal, September 17–21, 2015. The Association for Computational Linguistics, pp 379–389. https://doi.org/10.18653/v1/d15-1044

Sansone C, Sperlí G (2022) Legal information retrieval systems: state-of-the-art and open issues. Inf Syst 106:101967

Savelka J, Ashley KD (2021) Discovering explanatory sentences in legal case decisions using pre-trained language models. In: Moens M, Huang X, Specia L, Yih SW (eds) Findings of the association for computational linguistics: EMNLP 2021, Virtual Event/Punta Cana, Dominican Republic, 16–20 November, 2021. Association for Computational Linguistics, pp 4273–4283. https://doi.org/10.18653/v1/2021.findings-emnlp.361

See A, Liu PJ, Manning CD (2017) Get to the point: summarization with pointer-generator networks. In: Barzilay R, Kan M (eds) Proceedings of the 55th annual meeting of the association for computational linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, vol 1: Long Papers. Association for Computational Linguistics, pp 1073–1083. https://doi.org/10.18653/v1/P17-1099

Shen J, Pang R, Weiss RJ, Schuster M, Jaitly N, Yang Z, Chen Z, Zhang Y, Wang Y, Skerrv-Ryan R, Saurous RA, Agiomvrgiannakis Y, Wu Y (2018) Natural TTS synthesis by conditioning WaveNet on MEL spectrogram predictions. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 4779–4783. https://doi.org/10.1109/ICASSP.2018.8461368

Snyder D, Ghahremani P, Povey D, Garcia-Romero D, Carmiel Y, Khudanpur S (2016) Deep neural network-based speaker embeddings for end-to-end speaker verification. In: 2016 IEEE spoken language technology workshop (SLT), pp 165–170. https://doi.org/10.1109/SLT.2016.7846260

Şulea OM, Zampieri M, Vela M, van Genabith J (2017) Predicting the law area and decisions of French Supreme Court cases. In: Proceedings of the international conference recent advances in natural language processing, RANLP 2017, INCOMA Ltd., Varna, Bulgaria, pp 716–722. https://doi.org/10.26615/978-954-452-049-6_092

Sweeney L (1996) Replacing personally-identifying information in medical records, the Scrub system. In: Proceedings of the AMIA annual fall symposium. American Medical Informatics Association, pp 333–337

Tuggener D, von Däniken P, Peetz T, Cieliebak M (2020) LEDGAR: a large-scale multi-label corpus for text classification of legal provisions in contracts. In: Proceedings of the language resources and evaluation conference, european language resources association, pp 1228–1234

Tumminia J, Kuznecov A, Tsilerides S, Weinstein I, McFee B, Picheny M, Kaufman AR (2021) Diarization of legal proceedings. Identifying and transcribing judicial speech from recorded court audio. Preprint, arXiv:2104.01304

Uzuner Ö, Sibanda T, Luo Y, Szolovits P (2008) A de-identifier for medical discharge summaries. Artif Intell Med 42(1):13–35

Wellner B, Huyck M, Mardis S, Aberdeen J, Morgan A, Peshkin L, A Y, Hitzeman J, Hirschman J (2007) Rapidly retargetable approaches to de-identification in medical records. J Am Med Inform Assoc: JAMIA 14(5):564–573

Wilson RA (1962) Computer retrieval of case law. Southwest Law J 16:409–438

Wong KF, Wu M, Li W (2008) Extractive summarization using supervised and semi-supervised learning. In: Proceedings of the 22nd international conference on computational linguistics (Coling 2008), Coling 2008 Organizing Committee, Manchester, UK, pp 985–992. https://aclanthology.org/C08-1124

Xiao C, Zhong H, Guo Z, Tu C, Liu Z, Sun M, Zhang T, Han X, Hu Z, Wang H, et al. (2019) CAIL2019-SCM: a dataset of similar case matching in legal domain. Preprint, arXiv:191108962

Xiao C, Hu X, Liu Z, Tu C, Sun M (2021) Lawformer: a pre-trained language model for Chinese legal long documents. AI Open 2:79–84

Zhong H, Xiao C, Guo Z, Tu C, Liu Z, Sun M, Feng Y, Han X, Hu Z, Wang H, et al (2018) Overview of cail2018: legal judgment prediction competition. Preprint, arXiv:181005851

Zhong H, Xiao C, Tu C, Zhang T, Liu Z, Sun M (2020) How does nlp benefit legal system: a summary of legal artificial intelligence. In: Proceedings of the annual meeting of the association for computational linguistics, pp 5218–5230

Zhu X, Goldberg AB, Gael JV, Andrzejewski D (2007) Improving diversity in ranking using absorbing random walks. In: Sidner CL, Schultz T, Stone M, Zhai C (eds) Human language technology conference of the north american chapter of the association of computational linguistics, proceedings, April 22–27, 2007, Rochester, New York, USA. The Association for Computational Linguistics, pp 97–104. https://aclanthology.org/N07-1013/

Zhu K, Guo R, Hu W, Li Z, Li Y (2020) Legal judgment prediction based on multiclass information fusion. Complexity 2020:3089189. https://doi.org/10.1155/2020/3089189