

# Automatic Linking of Judgements to UK Supreme Court Hearings

**Hadeel Saadany**

Centre for Translation Studies  
University of Surrey  
United Kingdom  
hadeel.saadany@surrey.ac.uk

**Constantin Orăsan**

Centre for Translation Studies  
University of Surrey  
United Kingdom  
c.orasan@surrey.ac.uk

**Catherine Breslin**

Kingfisher Labs Ltd  
United Kingdom

**Sophie Walker**

Just Access  
United Kingdom

## Abstract

One of the most important archived legal material in the UK is the Supreme Court published judgements and video recordings of court sittings for the decided cases. The impact of Supreme Court published material extends far beyond the parties involved in any given case as it provides landmark rulings on arguable points of law of the greatest public and constitutional importance. However, the recordings of a case are usually very long which makes it both time and effort consuming for legal professionals to study the critical arguments in the legal deliberations. In this research, we summarise the second part of a combined research-industrial project for building an automated tool designed specifically to link segments in the text judgement to semantically relevant timespans in the videos of the hearings. The tool is employed as a User-Interface (UI) platform that provides a better access to justice by bookmarking the timespans in the videos which contributed to the final judgement of the case. We explain how we employ AI generative technology to retrieve the relevant links and show that the customisation of the GPT text embeddings to our dataset achieves the best accuracy for our automatic linking system.

## 1 Introduction

In the UK, HM Courts & Tribunals Service (HMCTS) publishes both live and recorded videos of court hearing sessions across different jurisdictions. This has been a tradition going on as back as 2008 for different types of courts in the UK. The main objective for this is to improve public access to, and understanding of, the work of the courts. The Supreme Court sits above all the UK separate jurisdictions and as the final court of appeal, its decisions contribute to the development of United Kingdom law. Moreover, the court

hearings crucially aid in new case preparation, provide guidance for court appeals, help in legal training and even guide future policy.

However, there are two main obstacles to making use of this rich material to learn more about the judicial system and have a better access to justice. First, the audio/video material for a case typically spans over several hours on several days, which makes it both time and effort consuming for legal professionals to extract important information relevant to their needs. Second, currently, the existing need for legal transcriptions, covering 449K cases p.a in the UK across all court tribunals, is largely met by human transcribers (Sturge, 2021). This, of course, makes the recorded material, which is rich of legal arguments relevant to how a judgement is reached, difficult to navigate either in text format or in its original audio-visual format.

In this research, we present a combined research-industrial effort to construct an integrated system for the automatic navigation of segments in the media data of UK Supreme Court hearings based on their semantic relevance to particular paragraph(s) in the text of the judgement issued following the hearing. Based on the timing metadata of the court hearing transcription segments, we manage to assign bookmarks on the video sessions and link them to their semantically relevant paragraphs in the judgement text. The main objective of the video bookmarking is to provide legal professionals, as well as the general public an automatic navigation tool that pins down the arguments and legal precedents presented in the long hearing sessions and which are of particular importance to how the judges made their decision on the case. We call our system the Judgement-to-Hearing Automatic Linking (J-HAL) and we deployed it as a User-Interface platform that can be used by legal professionals, academics and the

public.

Figure 1 shows a snapshot of the UI we created. On the left side of the screen, the paragraphs of the written judgement are displayed. The user can use the scroll down button to choose a specific paragraph in the judgement. On the right side, the timespan in the court hearing video that is semantically relevant to the legal point mentioned in the selected judgement paragraph is displayed along with temporal metadata (session number, day and time). The user can play the particular timespan and go back and forth around it as well as read our tool’s transcription of the speech. Major stakeholders from legal domain have shown interest in employing our UI as a bookmarking tool for identifying legally critical minutes in video sessions of Supreme court hearings to be used by legal academics and professionals. Also, our UI is currently in the application process to be published as an innovative patent by the UK Intellectual Property Office.

In this paper, we explain how we establish the automatic linking between judgement paragraphs and video session bookmarks through Information Retrieval (IR) models. First, in section 2, we briefly summarise relevant work for IR in the legal domain. Then, in section 3, we explain our system’s pipeline and its objectives. In section 4, we show how we compiled and preprocessed our dataset. Then, in section 5, we summarise our experiments on this dataset using a zero-shot IR approach as a pre-fetching stage and how we customise the GPT text embeddings for optimising our system. Finally, in section 6, we present our conclusion on the experiments conducted as well as our future plans for improving the linking system.

## 2 Related Work

Recently there has been an increased interest in employing NLP techniques to aid the textual processing in the legal domain (Elwany et al., 2019; Nay, 2021; Mumcuoğlu et al., 2021; Frankenreiter and Nyarko, 2022). The main focus has been on legal document summarisation (Shukla et al., 2022; Hellesoe, 2022), predicting judgements (Aletras et al., 2016; Trautmann et al., 2022) and contract preprocessing and generation (Hendrycks et al., 2021; Dixit et al., 2022). Moreover, NLP methods for Information Extraction and Textual Entailment have been extensively used in the domain of legal NLP to either find an answer to a legal query in

legal documents (Zheng et al., 2021) or to connect textual data (Rabelo et al., 2020). For example, Chalkidis et al. (2021) experiment with different IR models to extract relevant EU and UK legislative acts that are important for organisations’ *regulatory compliance* which they need to ensure it complies with the relevant laws. Their experiments show that fine-tuning a BERT model on an in-domain classification task is the best pre-fetcher for their dataset. Similarly, Kiyavitskaya et al. (2008) use textual semantic annotation to extract government regulations in different countries which companies and software developers are required to comply with. They show that AI-based IR tools are effective in reducing the human effort to derive requirements from regulations. Although there has recently been significant increase in the legal IR research, the processing and deployment of spoken court hearings for legal IR has not received the same attention as understanding and extracting information from textual legal data. In this research, we introduce an industrial product that employs IR tools to automatically connect judgements and videos of court hearings.

## 3 Linking Judgements and Case Hearings

The pipeline for automatically connecting judgements of decided cases to relevant bookmarks in the court hearing videos consists of two stages. First, we build a customised speech-to-text language model and employ NLP methods to improve the quality of the court hearing video transcriptions. The objective of this stage is to obtain a high quality transcript for our automatic retrieval model in the second stage. The second stage consists of building an IR system capable of extracting the best n-links between a paragraph(s) of the judgement text and the timespans of the transcribed video sessions of that particular case. The links are then translated into timestamp bookmarks in the long videos of each case to be used in constructing our UI. The first stage of the pipeline is beyond the scope of the current paper and is described in (Saadany et al., 2022, 2023). In the following sections, we explain the methodology we adopt for conducting the second stage in the system pipeline where the automatic linking is established between judgement text and video bookmarks.

In the second stage, we treat the linking of a judgement paragraph to the relevant timespan

1. Part III of the Matrimonial and Family Proceedings Act 1984 was enacted to give the English court the power to grant financial relief after a marriage had been dissolved (or annulled) in a foreign country. This appeal raises for the first time at this appellate level the proper approach to the operation of Part III of the 1984 Act.

2. Mr and Mrs Agbaje (“the husband” and “the wife”) were married for 38 years prior to their divorce in 2005 on the husband’s petition in Nigeria. They were born in Nigeria, but both have British and Nigerian citizenship. All five children of the family were born in England. The wife has been living in England continuously since 1999, when the marriage broke down. The assets are about £700,000, of which £530,000 represents two houses in London in the husband’s name, and the balance represents properties in Nigeria. The Nigerian court awarded the wife a life interest in a property in Lagos (which, as found by the Nigerian court, had a capital value of about £86,000) and a lump sum which was the equivalent of about £21,000.



00:31:50.440 - 00:32:39.870 - Day 1 Session 1

I'm going to come onto this later, when you look at the judgement of Lord Justice Ward, he has made the point, well, if you don't grant an anti-suit injunction, but you stay the English petition here on the basis that Nigeria is the appropriate forum. How can it be right to say that Nigeria shouldn't deal with the case? Well, the irony is that the judge who dealt with the application postulated the possibility of the wife making a Part 3 application in his judgement. I was dealing with the the forum shopping acquittal.

Figure 1: User-Interface for Linking Judgement to Bookmarks in Video Court Sessions

transcripts of a video session as a text retrieval task. In NLP, text retrieval is usually divided into two sub tasks according to the length of the text: query-to-passage retrieval and query-to-document retrieval (Xu et al., 2022). For our use case, it is a document-to-document retrieval task where we first transcribe the video sessions by a custom speech-to-text language model that we developed in stage one, and then segment the judgement into paragraphs where a paragraph(s) is treated as a query and the transcript of the case is the corpus in which we search for an answer to that query. It is important to note that the judgement text and the timespans of a transcript usually consist of several lines of text which makes the task more challenging than the traditional query-to-document IR where the query is typically short. Another challenge specific of our data is that we attempt to establish a link between texts that belong to two different language registers: written and spoken (Peters, 2003; Matthiessen and Halliday, 2009). Linguistically, the complexity of speech is “choreographic” (Halliday, 2007) where meta-linguistic elements such as intonation, loudness or quietness, pausing, stress, pitch range and gestures communicate semantic connotations. Moreover, spoken language is characterised by complex sentence structures with low lexical density (fewer high content words per clause), whereas written language typically contains simple sentence structures with high lexical density (more high content words per clause) (Halliday, 2007). Thus, the retrieval task in our case is nontraditional as it needs a careful preprocessing and segmentation of the spoken and the written datasets to obtain accurate results. In the following sections, we explain how we compiled and preprocessed our training dataset.

## 4 Data Compilation and Processing

For training our IR models, we extracted 7 case judgements consisting of 1.4M tokens scraped from the official site of the UK Supreme Court<sup>1</sup>. As for the transcription data, it consisted of 53 hours of video material for the selected cases obtained from the UK National Archive<sup>2</sup>. The video sessions were transcribed by our custom speech-to-text model which we trained in stage one of J-HAL system pipeline. We then ran a number of preprocessing steps to obtain the best linking accuracy between a judgement segment and the relevant timespans in the transcripts.

The main challenge in preprocessing the dataset was how to segment the judgement text into semantically cohesive sections that would be treated as queries in our IR method. We noticed that typically the Supreme Court judgement is structured manually into sections such as: “Introduction”, “The context”, “Facts of the Case”, “The Outcome of the Case”, etc. However, after we carefully scrutinised the dataset, we found that the naming of sections is not consistent. On the other hand, the judgement texts are consistently divided into enumerated paragraphs (typically a digit(s) followed by a dot). We opted, therefore, for segmenting the judgement text into windows of enumerated paragraphs. After experimenting with different window sizes, the optimum window size consisted of three enumerated paragraphs with an average length of 389 tokens per segment. As for the preprocessing of the transcription, it consisted mainly of excluding very short timespans since they were mostly either interjections (e.g. “Yes, sorry, I’m not following”, “I beg your pardon.”, etc.) or reference to logistics of the hearing (e.g. “This

<sup>1</sup><https://www.supremecourt.uk/decided-cases/>

<sup>2</sup><https://www.nationalarchives.gov.uk/>

is your paper, isn't it?", "Please turn to the next page. ", etc.). We chose to exclude transcription spans less than 50 tokens as an empirical threshold for semantically significant conversation units. For both the judgement and transcript data, we cleaned empty lines and extra spaces but kept punctuation intact as it is essential in identifying names of cases and legal provisions<sup>3</sup>.

## 5 Zero-shot Information Retrieval

The ability of an IR system to retrieve the top-N most relevant results is usually assessed by comparing its performance with human-generated similarity labels on a sentence-to-sentence or query-to-document similarity dataset(s) (e.g. Agirre et al., 2014; Boteva et al., 2016; Thakur et al., 2021). In order to create a human-generated evaluation dataset, we needed to assign human annotators to manually check the correct links between judgement segments and the timespans of video hearing transcripts for each of our chosen cases. However, in our use case, this is not feasible since to annotate one Supreme Court case with, for example, 50 judgement segments and 300 timespans of video transcript, the annotators will need to read 50 x 300 judgement-timespan link, which amounts to 15,000 doc-to-doc link per case. To overcome this problem, we adopted a zero-shot IR approach.

Thus to create a dataset for annotation, we first experimented with different ways to encode the judgement segments and transcription timespans as numeric vectors for a single case in our dataset. We used the cosine similarity as our semantic distance metric to extract the top closest 20 transcript timespans per judgement segment in vector space. Then, we assigned a human annotator, post-graduate law student, to evaluate the first 20 links produced by the different models. The annotator compared each judgement segment against each timespan to choose either 'Yes' there is a semantic link or 'No' there is not. The IR models used for our experiments are the following:

### A. Frequency-based Methods (keyword search)

**Okapi BM25** (Robertson et al., 2009): BM25 is a traditional keyword search based on a bag-of-words scoring function estimating the relevance

<sup>3</sup>The UK legal system has a unique punctuation style for case names such as "R v Chief Constable of South Wales [2020] EWCA Civ 1058" which are crucial in understanding legal precedents.

of a document  $d$  to a query  $q$ , based on the query terms appearing in  $d$ . It is a modified version of the tf-idf function where the ranking scores change based on the length of the document  $d$  in words, and the average  $d$  length in the corpus from which documents are drawn.

### B. Embedding-based Methods

**Document Similarity with Pooling:** We experimented with different pooling methods of the GloVe (Pennington et al., 2014a) pretrained word embeddings. The GloVe vector embeddings are created by unsupervised model training on general domain data (Pennington et al., 2014b). We create vectors for the judgement segment and the transcripts spans from the mean, minimum and maximum values of the GloVe embeddings.

**Entailment Search:** We use embeddings from a pretrained model for textual entailment which is trained to detect sentence pair relations, i.e. one sentence entails or contradicts the other. We employ the Microsoft MiniLM model (Wang et al., 2020) which is trained on the Microsoft dataset MiniLM-L6-H384-uncased and fine-tuned on a 1B sentence pairs dataset. The potential link in this case is whether or not the judgement paragraph(s) entails the particular segment of the video transcript.

**Legal BERT:** Our dataset comes from the legal domain which has distinct characteristics such as specialised vocabulary, particularly formal syntax, and semantics based on extensive domain-specific knowledge (Williams, 2007; Haigh, 2018). For this reason, we employed Legal BERT (Chalkidis et al., 2020) which is a family of BERT models for the legal domain pre-trained on 12 GB of diverse English legal text from several fields (e.g., legislation, court cases, contracts). The judgement text and the video transcript data were converted into the Legal BERT pretrained word embeddings.

**Asymmetric Semantic Search:** Asymmetric similarity search refers to finding similarity between unequal spans of text, which may be particularly applicable to our case where the judgement text may be shorter than the span of the video transcript. For this purpose, we created the embeddings using the MS MARCO model (Hofstätter et al., 2021) which is trained on a large scale IR corpus of 500k Bing query examples.

**GPT Question-answer linking:** In this setting a question-answer linking approach is adopted

Model	MAP@5	Recall@5	MAP@10	Recall@10	MAP@15	Recall@15
<b>GPT</b>	<b>0.96</b>	<b>0.33</b>	<b>0.89</b>	<b>0.57</b>	<b>0.85</b>	0.77
<b>Entailment</b>	0.87	0.32	0.85	0.55	0.82	<b>0.79</b>
<b>Glove</b>	0.81	0.27	0.77	0.53	0.61	0.78
<b>BM25</b>	0.87	0.29	0.81	0.53	0.78	0.77
<b>Asymmetric</b>	0.94	0.32	0.88	0.54	0.83	0.77
<b>LegalBert</b>	0.83	0.30	0.82	0.55	0.79	0.78

Table 1: Results of Unsupervised IR for Linking Judgements in One Case

Model	MAP@5	Recall@5	MAP@10	Recall@10	MAP@15	Recall@15
<b>GPT</b>	<b>0.691</b>	<b>0.391</b>	<b>0.622</b>	0.657	<b>0.711</b>	<b>0.914</b>
<b>BM25</b>	0.655	0.377	0.612	<b>0.659</b>	0.698	0.902
<b>Entailment</b>	0.615	0.348	0.568	0.611	0.66	0.885
<b>Glove</b>	0.526	0.316	0.506	0.602	0.607	0.884
<b>Asymmetric</b>	0.602	0.347	0.553	0.619	0.664	0.908
<b>LegalBert</b>	0.557	0.326	0.531	0.613	0.632	0.896

Table 2: Results of Unsupervised IR for linking Judgements in Entire Dataset

where the selected judgement text portion is treated as a question, and the segments of the video transcript as potential answers. We use pretrained embeddings obtained from OpenAI’s GPT latest text-embedding-ada-002 model to find answers in video timespans for each segment in the judgement which is treated as a prompt query.

To assess the performance of each model in comparison to the human judgement, we calculated the Mean Average Precision (MAP) which is the *de facto* IR metric:

$$MAP = \frac{1}{Q} \sum_{q=1}^Q AP(q) \quad (1)$$

where  $Q$  is the total number of queries, in our case the judgement segments, and  $AP(q)$  is the average precision of a single query  $q$ .  $AP(q)$  evaluates whether all of the timespans assigned as relevant by the annotator are ranked the highest by the model. We calculated MAP for the first 5, 10, and 15 judgement-timespan pairs.

As can be seen from Table 1, the GPT model demonstrated the best performance in comparison to the other models. Thus, to create a dataset for annotation for the rest of the cases, we extracted the top 15 links for each judgement-transcript segment according to the cosine similarity scores of the GPT embedding model. We also extracted 5 links with the lower ranks (50 to 55) to avoid bias to the GPT model and randomly shuffled the 20 links for each judgement-transcript segments.

After this processing, the dataset constructed for manual annotation consisted of 3620 judgement-to-transcript documents. The human annotators were again asked to judge whether the extracted timespan transcripts are semantically linked or not linked to the judgement paragraph(s). This was done with a specially designed interface which is similar to the UI presented in Figure 1.<sup>4</sup>

The human annotations were compared to the results of all the embedding models mentioned above. As shown in Table 2, the GPT text embedding model again shows superiority over the other models. Thus, the approach of treating the judgement segment as a query and the transcription of the video sessions as the corpus in which we try to find the answer gives the best MAP results for the first 5, 10 and 15 links. However, it should be pointed out that our use case is different than a typical IR task where the efficacy of the model is evaluated by its ability to get the best links in the very first few hits (optimally hits 1 to 5). In our case, the Recall, rather than the average precision, is of higher importance. The reason is that the output of the model is used to extract the transcription temporal metadata which is then used to bookmark the video sessions at the relevant parts where the UI user can watch or draw the cursor around the bookmark to get more information. Accordingly, our system’s priority is to extract as

<sup>4</sup>The code for the annotation interface is available at <https://github.com/dinel/SimilarityLinksAnnotator>.

many relevant bookmarks as possible from all the true relevant links in the long video sessions so that the user can understand how the judgement point is argued in the session.

Table 2 also reveals that the BM25’s recall performance is on par with the GPT specifically in retrieving relevant links in the top 10 and 15 hits. Although compute-intensive approaches generalise better, they are significantly slow as compared to BM25 which is a lightweight search model. In building our system, we find that the GPT approach is a midway solution because of its superior performance and speedy API calls for extracting embeddings.

### 5.1 Model Optimisation

To optimise the performance of the best IR model, we customised the GPT embeddings to be more domain specific. The GPT embedding model used for our retrieval is the text-embedding-ada-002, which was introduced by OpenAI in December 2022 as their state-of-the-art text embedding model. It is trained on different datasets used for text search, text similarity, and code search. In order to customise the GPT embeddings, we follow the OpenAI method for embedding customisation (Sanders, 2023). Thus, we train a classification model on our humanly-annotated data with the following objective:

$$SE_{\min} = \min SE(x) \mid x \in \{-1, -0.99, \dots, 1\} \quad (2)$$

where  $x$  is the cosine similarity threshold between the positive and negative class which we obtain by sweeping between cosine similarity scores from -1 to 1 in steps of 0.01 till we get the lowest standard error of mean  $SE_{\min}$  for the cosine similarity distribution. The output of this training is a matrix  $M$  that we multiply by the embedding vector  $v$  of each judgement and transcript segment. This multiplication produces customised embeddings which are more adapted to our legal dataset. The graphs in Figures 2 and 3 show that the overlap between the distribution of the cosine similarities for relevant and irrelevant judgement-hearing links improves from  $70.5\% \pm 2.7\%$  with the original GPT embeddings to  $73.0\% \pm 2.6\%$  with the customised embeddings. The customised embeddings contribute to a more accurate automatic linking between judgement text and video court hearings. An example of the GPT

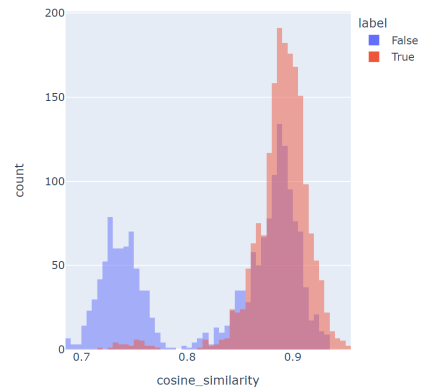


Figure 2: Cosine Similarity Distribution with Original GPT Embeddings

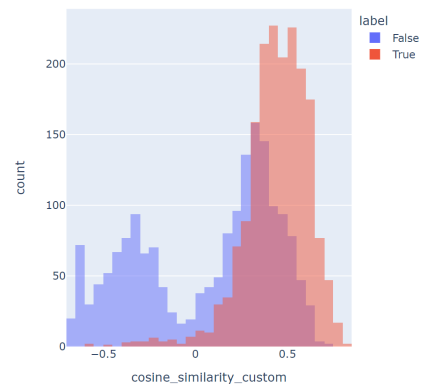


Figure 3: Cosine Similarity Distribution with Customised GPT Embeddings

retrieval output which we use as the back-end model for our UI is shown in Appendix A. As per our human annotator markings, each text colour indicates a legal point presented in the judgement text and its semantically relevant argument in the hearing transcript is presented in the same colour.

## 6 Conclusion

This research presented the second stage of our pipeline which employs generative AI to automatically link a judgement text of a decided case in the UK Supreme Court to its video hearings. The IR system we provide assists users in extracting the arguments and information they may find useful in understanding the particular case they are studying. The system does not, however, explicitly return answers to questions legal professionals may have on a legal precedent. The UI we provide supports the users in browsing or filtering the lengthy videos of the court hearing sessions by searching through hundreds of video timespans and then provides a set of need-to-watch bookmarks that are crucial in understanding the

judgement decided for the case. The implications of this tool extends beyond providing practical aid for legal professionals and academics to expanding the public’s ability to access information on court proceedings and justice in general. In the future stages of the project, we aim to expand our annotated linking dataset and explore the effectiveness of coupling judgements and video hearings according to common legal entities such as articles, legal provisions and names of similar cases. We also aim to adopt a similar methodology for constructing similar UI tools for other purpose domains such as bookmarking recorded lecture spans to educational text books or linking written documents to recorded meetings in the business sector.

## Acknowledgements

We would like to thank and acknowledge the effort exerted by our legal expert team of annotators who took the time to carefully read the dataset and provide relevancy labels.

## References

- Eneko Agirre, Carmen Banea, Claire Cardie, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, Weiwei Guo, Rada Mihalcea, German Rigau, and Janyce Wiebe. 2014. Semeval-2014 task 10: Multilingual semantic textual similarity. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*, pages 81–91.
- Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel PreoŃuc-Pietro, and Vasileios Lampos. 2016. Predicting judicial decisions of the european court of human rights: A natural language processing perspective. *PeerJ computer science*, 2:e93.
- Vera Boteva, Demian Gholipour, Artem Sokolov, and Stefan Riezler. 2016. A full-text learning to rank dataset for medical information retrieval. In *Proceedings of the 38th European Conference on Information Retrieval*. <http://www.cl.uni-heidelberg.de/~riezler/publications/papers/ECIR2016.pdf>.
- Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. **LEGAL-BERT: The muppets straight out of law school**. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2898–2904, Online. Association for Computational Linguistics.
- Ilias Chalkidis, Manos Fergadiotis, Nikolaos Manginas, Eva Katakalous, and Prodromos Malakasiotis. 2021. Regulatory compliance through doc2doc information retrieval: A case study in eu/uk legislation where text similarity has limitations. *arXiv preprint arXiv:2101.10726*.
- Abhishek Dixit, Vipin Deval, Vimal Dwivedi, Alex Norta, and Dirk Draheim. 2022. Towards user-centered and legally relevant smart-contract development: A systematic literature review. *Journal of Industrial Information Integration*, 26:100314.
- Emad Elwany, Dave Moore, and Gaurav Oberoi. 2019. Bert goes to law school: Quantifying the competitive advantage of access to large legal corpora in contract understanding. *arXiv preprint arXiv:1911.00473*.
- Jens Frankenreiter and Julian Nyarko. 2022. Natural language processing in legal tech. *Legal Tech and the Future of Civil Justice (David Engstrom ed.)*.
- Rupert Haigh. 2018. *Legal English*. Routledge.
- Michael Alexander Kirkwood Halliday. 2007. *Language and Education: Volume 9*. A&C Black.
- Lui Joseph Hellesoe. 2022. *Automatic Domain-Specific Text Summarisation With Deep Learning Approaches*. Ph.D. thesis, Auckland University of Technology.
- Dan Hendrycks, Collin Burns, Anya Chen, and Spencer Ball. 2021. Cuad: An expert-annotated nlp dataset for legal contract review. *arXiv preprint arXiv:2103.06268*.
- Sebastian Hofstätter, Sheng-Chieh Lin, Jheng-Hong Yang, Jimmy Lin, and Allan Hanbury. 2021. Efficiently Teaching an Effective Dense Retriever with Balanced Topic Aware Sampling. In *Proc. of SIGIR*.
- Nadzeja Kiyavitskaya, Nicola Zeni, Travis D Breaux, Annie I Antón, James R Cordy, Luisa Mich, and John Mylopoulos. 2008. Automating the extraction of rights and obligations for regulatory compliance. In *Conceptual Modeling-ER 2008: 27th International Conference on Conceptual Modeling, Barcelona, Spain, October 20-24, 2008. Proceedings 27*, pages 154–168. Springer.
- Christian MIM Matthiessen and Michael Alexander Kirkwood Halliday. 2009. Systemic functional grammar: A first step into the theory.
- Emre Mumcuođlu, Ceyhun E Öztürk, Haldun M Ozaktas, and Aykut Koç. 2021. Natural language processing in law: Prediction of outcomes in the higher courts of turkey. *Information Processing & Management*, 58(5):102684.
- John J. Nay. 2021. *Natural Language Processing for Legal Texts, DOI=10.1017/9781316529683.011*, page 99–113. Cambridge University Press.
- Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014a. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

- Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014b. Glove: Global vectors for word representation. Figshare <https://nlp.stanford.edu/projects/glove/>.
- J Melanie Peters. 2003. *The Impact of Tele-Advice on the Community Nurses' Management of Leg Ulcers*. University of South Wales (United Kingdom).
- J Rabelo, MY Kim, R Goebel, M Yoshioka, Y Kano, and K Satoh. 2020. Coliee 2020: Methods for legal document retrieval and entailment, 2020. URL: [https://sites.ualberta.ca/~rabelo/COLIEE2021/COLIEE\\_2020\\_summary.pdf](https://sites.ualberta.ca/~rabelo/COLIEE2021/COLIEE_2020_summary.pdf).
- Stephen Robertson, Hugo Zaragoza, et al. 2009. The probabilistic relevance framework: Bm25 and beyond. *Foundations and Trends® in Information Retrieval*, 3(4):333–389.
- Hadeel Saadany, Catherine Breslin, Constantin Orăsan, and Sophie Walker. 2023. **Better transcription of uk supreme court hearings**. In *Workshop on Artificial Intelligence for Access to Justice (AI4AJ 2023)*.
- Hadeel Saadany, Constantin Orăsan, and Catherine Breslin. 2022. Better transcription of uk supreme court hearings. *arXiv preprint arXiv:2211.17094*.
- Ted Sanders. 2023. Customizing embeddings. OpenAI [https://github.com/openai/openai-cookbook/blob/main/examples/Customizing\\_embeddings.ipynb](https://github.com/openai/openai-cookbook/blob/main/examples/Customizing_embeddings.ipynb).
- Abhay Shukla, Paheli Bhattacharya, Soham Poddar, Rajdeep Mukherjee, Kripabandhu Ghosh, Pawan Goyal, and Saptarshi Ghosh. 2022. Legal case document summarization: Extractive and abstractive methods and their evaluation. *arXiv preprint arXiv:2210.07544*.
- Georgina Sturge. 2021. **Court statistics for England and Wales**. Technical report, House of Commons Library.
- Nandan Thakur, Nils Reimers, Andreas Rücklé, Abhishek Srivastava, and Iryna Gurevych. 2021. BEIR: A heterogeneous benchmark for zero-shot evaluation of information retrieval models. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*. <https://openreview.net/forum?id=wCu6T5xFjeJ>.
- Dietrich Trautmann, Alina Petrova, and Frank Schilder. 2022. Legal prompt engineering for multilingual legal judgement prediction. *arXiv preprint arXiv:2212.02199*.
- Wenhui Wang, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. 2020. **Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers**.
- Christopher Williams. 2007. *Tradition and change in legal English: Verbal constructions in prescriptive texts*, volume 20. Peter Lang.
- Guangwei Xu, Yangzhao Zhang, Longhui Zhang, Dingkun Long, Pengjun Xie, and Ruijie Guo. 2022. Hybrid retrieval and multi-stage text ranking solution at trec 2022 deep learning track. *TREC 2022 Deep Learning Track*.
- Lucia Zheng, Neel Guha, Brandon R Anderson, Peter Henderson, and Daniel E Ho. 2021. When does pretraining help? assessing self-supervised learning for law and the casehold dataset of 53,000+ legal holdings. In *Proceedings of the eighteenth international conference on artificial intelligence and law*, pages 159–168.



# Appendix A An Example of the Automatic Linking of Judgement Segment and Transcription Segments by GPT Embeddings

- Judgement:**

**Legal Point** → The complaint, very moderately advanced by Mr. Geekie QC, is that a "starting point" of undesirability, placing the burden upon the person wishing to cross-examine a child to show some "particular justification" for doing so, gives insufficient weight to the **Convention rights** of all concerned. All the parties in care proceedings are entitled to a fair hearing in the determination of their civil rights and obligations the parents who stand to lose their children if allegations of abuse are made out, the children who stand to lose their parents if allegations of abuse are made out, but also stand to suffer abuse or further abuse if they are left at home because those allegations cannot be proved. And it is not only their **article 6 rights** which are in play. The **civil rights** in issue are also **Convention rights** in themselves the right to respect for the family lives of the parents and their children but also the right to respect for the private lives of the children, which include their rights to be protected from attacks upon their physical and psychological integrity. Even a "stranger" child, whose future is not in issue in the proceedings
- Timestamps:**

**00:01:31.440,00:03:19.240,spk\_2 (The Judge),**

**Correct link** → Looking at what we say should be the starting point. Can I just consider, the issue of the welfare of the child and in the context and this feeds directly into my paragraph 47 where I say the welfare, of the child is irrelevant but not paramount consideration it links into Mr. Geekie's point about whether a non subject child, so a child who is simply a witness within proceedings has Article 8 rights engaged, so not as confident as Mr. Geekie on that. But what I would say is that if the court is considering a child as a witness and in no other capacity, we would say that that **child's welfare is relevant** and that probably actually answers the Article 8 point if that's engaged. But when we say welfare of a child is relevant, we mean for **any child witness, whether a party or no**. The way in which we had formulated the factors to be put into the balancing exercise came as is clear from paragraph 45 of the Medway case from the judgement of the Lady Justice Smith, where she went on to look at the various factors, she's put them in a much more generalized form.
- 00:22:40.710,00:24:57.600,spk\_1 (The lawyer),**

**Correct link** → An equal starting point for all competing interests and the -- of course, one can't begin to list all competing interests that might arise, but judges can be cautioned or advised to be alert to **the Article 6 and 8 rights of witnesses**, and in the way that My Lady, we've canvassed already, that would apply to different -- witnesses in the three different positions before the Court. **The Article 6 and 8 rights of parents, and the Article 6 and 8 rights of siblings. And if one sets those as separate categories as the appropriate path to begin thinking about, that at least gives a decision-making judge in an individual case the right ambit of decisions to consider. The present authorities rely significantly upon ....**